Rare copy number variants in *NRXN1* and *CNTN6* increase risk for Tourette syndrome

Alden Y. Huang[1,2,3], Dongmei Yu[4,5], Lea K. Davis[6], Jae Hoon Sul[1,2], Fotis Tsetsos[7], Vasily Ramensky[1,2,8], Ivette Zelaya[1-3], Eliana Marisa Ramos[1,2], Lisa Osiecki[4], Jason A. Chen[1-3], Lauren M. McGrath[9], Cornelia Illmann[2], Paul Sandor[10], Cathy L. Barr[11], Marco Grados[12], Harvey S. Singer[12], Markus M. Nöthen[13,14], Johannes Hebebrand[15], Robert A. King[16], Yves Dion[17], Guy Rouleau[18], Cathy L. Budman[19], Christel Depienne[20,21], Yulia Worbe[21], Andreas Hartmann[21], Kirsten R. Muller-Vahl[22], Manfred Stuhrmann[23], Harald Aschauer[24,25], Mara Stamenkovic[24], Monika Schloegelhofer[25], Anastasios Konstantinidis[24,26], Gholson J. Lyon[27], William M. McMahon[28], Csaba Barta[29], Zsanett Tarnok[30], Peter Nagy[30], James R. Batterson[31], Renata Rizzo[32], Danielle C. Cath[33,34], Tomasz Wolanczyk[35], Cheston Berlin[36], Irene A. Malaty[37], Michael S. Okun[37], Douglas W. Woods[38,39], Elliott Rees[40], Carlos N. Pato[41], Michele T. Pato[41], James A Knowles[42], Danielle Posthuma[43], David L. Pauls[4], Nancy J. Cox[6,7], Benjamin M. Neale[4,5,44], Nelson B. Freimer[1,2], Peristera Paschou[6,45,49], Carol A. Mathews[47,49], Jeremiah M. Scharf[4,5,47,48,49,*,‡], & Giovanni Coppola[1,2,49,*] on behalf of the Tourette Syndrome Association International Consortium for Genomics (TSAICG) and the Gilles de la Tourette Syndrome GWAS Replication Initiative (GGRI).

[1]Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA
[2]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA
[3]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA
[4]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[6]Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
[7]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA
[8]Moscow Institute of Physics and Technology, Dolgoprudny, Institusky 9, Moscow, Russian Federation
[9]Department of Psychology, University of Denver, Denver, CO, USA
[10]Toronto Western Hospital, University Health Network and Youthdale Treatment Centres, University of Toronto, Toronto, Canada
[11]Krembil Research Institute, University Health Network, Toronto, Canada
[12]Johns Hopkins University School of Medicine, Baltimore, MD, USA

[13]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany
[14]Institute of Human Genetics, University of Bonn, Bonn, Germany
[15]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
[16]Yale Child Study Center, Yale University School of Medicine, New Haven, CT, USA
[17]University of Montréal , Montreal, Canada
[18]Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Canada
[19] Hofstra Northwell School of Medicine, Hempstead, NY, USA
[20]Département de Médicine Translationnelle et Neurogénétique, IGBMC, CNRS UMR 7104/INSERM U964/Université de Strasbourg,Illkirch, France
[21]Brain and Spine Institute, UPMC/INSERM UMR_S1127, Paris, France
[22]Clinic of Psychiatry, Social Psychiatry and Psychotherapy, Hannover Medical School, Hannover, Germany
[23]Institute of Human Genetics, Hannover Medical School, Hannover, Germany
[24]Medical University Vienna, Department of Psychiatry and Psychotherapy, Vienna, Austria.
[25]Biopsychosocial Corporation, Vienna, Austria
[26]Center for Mental Health Muldenstrasse, BBRZMed, Linz, Austria
[27]Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA
[28]Department of Psychiatry, University of Utah, UT, USA
[29]Institute of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary
[30]Vadaskert Child and Adolescent Psychiatric Hospital, Budapest, Hungary
[31]Children's Mercy Hospital, Kansas City, KS, USA
[32]Dipartimento di Medicina Clinica e Sperimentale, Università di Catania, Catania, Italy
[33]Department of Psychiatry, University Medical Center Groningen & Drenthe Mental Health Center, Groningen, the Netherlands
[34]Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands
[35]Department of Child Psychiatry, Medical University of Warsaw, Warsaw, Poland
[36]Penn State University College of Medicine, Hershey, PA, USA
[37]Department of Neurology and Center for Movement Disorders and Neurorestoration, University of Florida, Gainesville, FL, USA
[38]Marquette University, Milwaukee, WI, USA
[39]University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[40]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, Wales, United Kingdom
[41]SUNY Downstate Medical Center, Brooklyn, NY, USA
[42]Department of Psychiatry & Behavioral Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[43]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, VU University Amsterdam, Amsterdam, the Netherlands
[44]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
[45]Department of Molecular Biology and Genetics, Democritus University of Thrace, Greece
[46]Department of Psychiatry, Genetics Institute, University of Florida, Gainesville, FL, USA
[47]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA
[48]Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA

[49]These authors jointly directed this work.
[‡]Lead Contact
*Correspondence: jscharf@partners.org (J.M.S.), gcoppola@ucla.edu (G.C.).

Members of the Tourette Syndrome Association International Consortium for Genetics include:

Cathy L. Barr, James R. Batterson, Cheston Berlin, Ruth D. Bruun, Cathy L. Budman, Danielle C. Cath, Sylvain Chouinard, Giovanni Coppola, Nancy J. Cox, Sabrina Darrow, Lea K. Davis, Yves Dion, Nelson B. Freimer, Marco A. Grados, Matthew E. Hirschtritt, Alden Y. Huang, Cornelia Illmann, Robert A. King, Roger Kurlan, James F. Leckman, Gholson J. Lyon, Irene A. Malaty, Carol A. Mathews, William M. MaMahon, Benjamin M. Neale, Michael S. Okun, Lisa Osiecki, David L. Pauls, Danielle Posthuma, Vasily Ramensky, Mary M. Robertson, Guy A. Rouleau, Paul Sandor, Jeremiah M. Scharf, Harvey S. Singer, Jan Smit, Jae-Hoon Sul, Dongmei Yu

Members of the Gilles de la Tourette Syndrome Genome-wide Association Study Replication Initiative include:

Harald Aschauer, Csaba Barta, Danielle C. Cath, Christel Depienne, Andreas Hartmann, Johannes Hebebrand, Anastasios Konstantinidis, Kirsten R. Muller-Vahl, Peter Nagy, Markus M. Nöthen, Peristera Paschou, Renata Rizzo, Guy Rouleau, Monika Schlögelhofer, Mara Stamenkovic, Manfred Stuhrmann, Fotis Tsetsos, Zsanett Tarnok, Tomasz Wolanczyk, Yulia Worbe

SUMMARY

Tourette syndrome (TS) is a model neuropsychiatric disorder thought to arise from abnormal development and/or maintenance of cortico-striato-thalamo-cortical circuits. TS is highly heritable, but its underlying genetic causes are still elusive, and no genome-wide significant loci have been discovered to date. We analyzed a European ancestry sample of 2,434 TS cases and 4,093 ancestry-matched controls for rare (<1% frequency) copy-number variants (CNVs) using SNP microarray data. We observed a global enrichment of TS-associated CNVs that was most prominent in large (>1 Mb), singleton events (OR=3.0, 95%CI [1.4-6.6], $p=5.0x10^{-3}$). We also identified two individual, genome-wide significant loci, each conferring a substantial increase in TS risk (*NRXN1* deletions, OR=20.3, 95%CI [2.6-156.2]; *CNTN6* duplications, OR=10.1, 95% CI [2.3-45.4]). Approximately 1% of TS cases carry one of these CNVs, indicating that rare structural variation contributes significantly to the genetic architecture of TS.

INTRODUCTION

Tourette syndrome (TS) is a complex neuropsychiatric disorder characterized by multiple chronic involuntary motor and vocal tics, with an estimated population prevalence of 0.3-0.9% (Scharf et al., 2015). Tics typically emerge during childhood and peak in adolescence, with a subsequent reduction in symptoms, supporting the notion that TS is neurodevelopmental in origin (Robertson et al., 2017). Most TS patients (>85%) present with additional neuropsychiatric co-morbidities, typically attention deficit hyperactivity disorder (ADHD) and obsessive-compulsive disorder (OCD) (Hirschtritt et al., 2015), although the risk for mood, anxiety, major depressive, and autism spectrum disorders (ASD) is also elevated (Burd et al., 2009; Hirschtritt et al., 2015). Consequently, TS is often considered a model neuropsychiatric disorder in that identification of its underlying molecular, cellular, and neurophysiologic etiology may be broadly applicable to a wide range of psychiatric disorders.

Neuroimaging (Greene et al., 2016; Marsh et al., 2009) and neurophysiology (Draper et al., 2014; Gilbert et al., 2004) studies suggest that TS and its associated comorbidities (e.g., OCD and ADHD) arise from dysregulated development and/or maintenance of parallel cortico-striatal-thalamo-cortical (CSTC) motor, limbic, and cognitive circuits (Jahanshahi et al., 2015). Though non-genetic factors have been associated with increased TS risk (Browne et al., 2016; Leivonen et al., 2016), TS is primarily a genetic disorder. Family studies indicate that children of affected parents have a 60-fold higher risk of developing TS or chronic tics (CT), a closely related disorder, compared to the general population (Browne et al., 2015). TS heritability is estimated to be 0.77 (Mataix-Cols et al., 2015), making it one of the most heritable complex neuropsychiatric disorders. Despite this strong genetic component, the identification of *bona-fide* TS susceptibility genes has proven challenging. Although linkage analyses have identified several candidate regions, there is little consensus across studies, suggesting that, as with other neuropsychiatric disorders, TS is genetically complex and heterogeneous (Robertson et al.,

2017). Similarly, analyses of TS genetic architecture using aggregated SNP data demonstrates that TS is highly polygenic, with the majority of inherited TS risk distributed throughout the genome (Davis et al., 2013), though an initial genome-wide association study (GWAS) did not yield any genome-wide significant loci, likely due to small sample size (Scharf et al., 2013).

Studies examining rare structural variation in individuals with TS have implicated several neurodevelopmental genes involved in neurite outgrowth and axonal migration. Rare chromosomal abnormalities affecting *CNTNAP2* (Verkerk et al., 2003) and *SLITRK1* (Abelson et al., 2005) have been found in isolated TS families, and exonic copy-number variants (CNVs) in *NRXN1* are reported in small genome-wide studies (Nag et al., 2013; Sundaram et al., 2010), though no locus has yet survived genome-wide correction for multiple testing. Because of the evidence suggesting that rare CNVs may have a role in TS etiology (Fernandez et al., 2012; McGrath et al., 2014), and since such variants contribute to susceptibility for other heritable neurodevelopmental disorders (NDDs) (Malhotra and Sebat, 2012), we assessed the impact of rare CNVs on TS disease risk in a large sample of 6,527 unrelated individuals of European ancestry. We demonstrate a global increase in the burden of large, rare CNVs in TS cases compared to controls driven primarily by large, singleton events, in particular large (>1Mb) deletions, consistent with marked genetic heterogeneity. We also report the first two TS susceptibility loci that meet genome-wide significance: deletions in *NRXN1* and duplications in *CNTN6*. Each confers a substantial increase in disease risk and together are present in 1% of TS cases.

RESULTS

An overview of the sample selection, quality control, CNV detection, and data analysis performed in this study is presented in Figure 1 and described in detail in the STAR methods. All TS cases and

controls were recruited through the Tourette Syndrome Association International Consortium for Genetics (TSAICG) or through the Gilles de la Tourette Syndrome GWAS Replication Initiative (GGRI), with additional controls selected from external studies. All DNA samples were genotyped on the Illumina OmniExpress SNP array platform (Table S1A). We restricted analysis to SNP assays common to all array versions. We conducted extensive quality control analyses including both SNP-based and CNV-based exclusion of outliers (Table S1B and STAR Methods) and genotype-based determination of ancestry (Figure S1). The final dataset consisted of 6,527 unrelated European ancestry samples: 2,434 individuals diagnosed with TS and 4,093 unselected controls.

Genome-wide detection of CNVs was performed using the consensus of two widely-used Hidden Markov Model (HMM)-based methods (STAR Methods). Additionally, we used a locus-specific, intensity-based clustering method to generate CNV genotypes in all samples across 11 common HapMap3 loci (Figure S2, Tables S2) for sensitivity analysis. Using the proportion of concordant HMM-based calls at these loci as a sensitivity measure, we confirmed the absence of any bias in CNV detection between cases and controls across all loci ($p=0.54$, Fisher's Exact test, Tables S3A and S3B) and between individuals ($p=0.15$, Welch's *t*-test, Table S3C, and STAR Methods). Post-call cleaning was performed (STAR Methods), and CNVs were annotated for genic content and frequency. CNVs were considered "genic" if they overlapped the exon of a known protein-coding Refseq transcript. Frequencies were defined based on a 50% overlap with other CNVs as described (CNV and SCZ Working Groups of the PGC, 2017); "singletons" denote CNVs with a frequency of one across the entire dataset. We filtered calls for rare (frequency < 1% or < 65 events) CNVs ≥30kb in length and spanning at least 10 probes. Finally, using a heuristically derived series of *in silico* validation metrics, we removed aberrant CNV calls due to mosaicism and misclassified rare events (Figure S4). In total, we resolved 9,375 rare CNV calls (Table S4).

Global burden analysis of rare CNVs in TS

An increase in rare CNV burden has been consistently demonstrated in other NDDs (CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium and Psychosis Endophenotypes International Consortium, 2017). To control for potential confounders, burden analysis was performed using multiple logistic regression (STAR Methods) and three different standardized burden metrics: 1) total number of CNVs (CNV count), 2) total genomic size of all CNVs (CNV length), and 3) number of genes affected (CNV gene count). For genic CNVs (n=4,604), we observed a modest but significantly increased CNV burden across all metrics (Figure 2A): CNV count (OR 1.05 [1.01-1.10], p=0.027), CNV gene count (OR 1.09 [1.01-1.17], p=0.019), and CNV length (OR 1.15 [1.07-1.24], $p=1.9x10^{-4}$). By contrast, no enrichment was seen in a comparable number (n=4,771) of non-genic events. The increased burden in TS was most significant for CNV length and consistent across each control set individually (Figure S4). To explore the CNV length burden further, we partitioned the data across a range of CNV size and frequency bins and observed the enrichment was mainly attributable to large (>1Mb; OR 1.26 [1.08-1.49], $p=5.3x10^{-3}$) (Figure 2B) and/or singleton CNVs (OR 1.13 [1.04-1.24], $p=2.9x10^{-3}$) (Figure 2C).

Enrichment of large, singleton events and clinically relevant CNVs

We next explored whether specific CNV classes were enriched in TS. Since the elevated TS CNV burden was confined to large and/or very rare events, we re-evaluated the CNV count burden restricted to singletons, stratified by CNV size. We observed a significant enrichment of singletons >500kb (OR 1.43 [1.07-1.93], p=0.010) that was further increased in the largest size category (>1Mb, OR 2.28 [1.39-3.79], $p=1.2x10^{-3}$), with greater enrichment for deletions >1Mb (OR 2.75 [1.28-5.23], $p=6.5x10^{-3}$) than

duplications (OR 1.98 [1.04-3.83], p=0.038) (Figure 3A). Notably, this enrichment was driven by CNVs spanning genes under strong evolutionary constraint (probability of Loss-of-Function Intolerance (pLI) score>0.9; Lek et al., 2015) (RR=2.65 [1.40-5.00], $p=2.7x10^{-3}$; Poisson regression controlling for sex, CNV quality score, and ancestry principal components; STAR Methods).

It is well established that certain regions of the human genome are prone to large, rare, recurrent CNVs associated with a broad range of NDDs (Malhotra and Sebat, 2012). To characterize the extent to which these known pathogenic CNVs might also confer risk for TS, we classified all rare CNV calls by clinical relevance according to American College of Medical Genetics (ACMG) guidelines (Kearney et al., 2011) and assessed for enrichment between cases and controls. Known pathogenic CNVs were identified in 1.9% of TS cases vs. 0.8% of controls (OR 3.03 [1.85-5.07], $p=1.5x10^{-5}$) (Figure 3B). Consistent with an increased pathogenicity of deletions compared to duplications, this enrichment was greater for deletions alone (OR per CNV 3.94 [1.83-8.95], $p=6.3 x10^{-4}$). By contrast, no increase in burden was observed among CNVs classified by the ACMG as either benign or of unknown significance.

Deletions in *NRXN1* and duplications in *CNTN6* confer substantial risk for TS

To test our sample for enrichment of rare CNVs at individual genomic loci, we conducted an unbiased, point-wise (segmental) genome-wide association test, treating deletions and duplications independently (STAR Methods). As non-overlapping CNVs affecting the same gene would be unaccounted for by segmental assessments of enrichment, we also performed a complementary gene-based test, conditioned on CNVs affecting exons. In contrast to SNP-based association studies, there is no established p-value threshold to indicate genome-wide significance for CNVs, as the number of rare CNV breakpoints per genome varies across individuals and detection platforms. Therefore, for both tests,

we established both locus-specific p-values (P_{seg} and P_{gene} for segmental and gene-based tests, respectively) and genome-wide corrected (P_{corr}) p-values empirically through 1,000,000 label-swapping permutations, using the max(T) method (Westfall and Troendle, 2008) to control for familywise error rate (FWER). Both tests converged on the same two loci, one for deletions and another for duplications, which were enriched among TS cases and survived genome-wide correction for multiple testing.

For deletions, the peak segmental association signal was located on chromosome 2p16 ($P_{seg}=7.0x10^{-6}$; $P_{seg\text{-}corr}=1.0x10^{-3}$; Figure 4A), corresponding to heterozygous losses across the first two exons of *NRXN1*, and found exclusively among TS cases (N=10, Figure 4B). In the gene-based test of exonic CNVs, heterozygous *NRXN1* deletions were also the most significant association genome-wide ($P_{gene}=5.9x10^{-5}$; $P_{gene\text{-}corr}=8.5x10^{-4}$), representing 12 cases (0.49%) and one control (0.02%), and corresponding to a substantially increased TS risk (OR 20.3 [2.6-156.2]). Consistent with previously identified pathogenic *NRXN1* deletions in ASD, SCZ, and epilepsy, these exon-spanning CNVs clustered at the 5' end of *NRXN1* and predominantly affected the *NRXN1*-α isoform (Ching et al., 2010).

The segmental association test for CNV duplications identified one genome-wide significant locus on chromosome 3p26 within *CNTN6* ($P_{seg}=5.4x10^{-5}$, $P_{seg\text{-}corr}=6.9x10^{-3}$) with a secondary peak located directly upstream ($P_{seg}=5.9x10^{-5}$,$P_{seg\text{-}corr}=6.9x10^{-3}$, Figures 4A and 4C). Closer inspection revealed an enrichment of large duplications spanning this gene. The gene-based test identified the same locus, exonic *CNTN6* duplications, with heterozygous gains found in 12 cases (0.49%) and 2 controls (0.05%), corresponding to an OR=10.1 [2.3-45.4] ($P_{gene}=2.5x10^{-4}$, $P_{gene\text{-}corr}=8.3x10^{-3}$). Notably, the *CNTN6* duplications in TS cases were considerably larger than those in controls (641 vs. 143 kb). 9 of 12 TS carriers harbored a duplication >500 kb in length, while *CNTN6* duplications in controls were <200kb.

We verified all genic CNV calls across *NRXN1* and *CNTN6* by inspection of probe-level intensity plots (Figures S5 and S6). No additional loci were significant after controlling for FWER, under either

segmental or gene-based tests of association, and we obtained similar results after pair-matching each case with its closest ancestrally matched control, suggesting that these results are not due to inter-European population stratification (STAR Methods and Figure S7).

DISCUSSION

In this study, we demonstrate a significant role for rare structural variation in the pathogenesis of TS, a still poorly understood neurodevelopmental disorder. We observe an increased global burden of rare CNVs and report two definitive TS risk loci that surpass empirical thresholds for genome-wide significance, deletions in *NRXN1* and duplications in *CNTN6*.

NRXN1 is a highly-studied, pre-synaptic cell-adhesion molecule involved in synaptogenesis and synaptic transmission at both glutamatergic and GABAergic synapses (Pak et al., 2015). The *NRXN1* gene is primarily transcribed from two alternative promoters, resulting in a full-length *NRXN1-α* isoform and a short C-terminal *NRXN1-β* isoform (Ushkaryov et al., 1992). *NRXN1-α* contains six alternative splice sites which, in combination, generate hundreds of unique transcripts that segregate within specific brain regions and cell types (Fuccillo et al., 2015; Schreiner et al., 2014). NRXN1-α isoforms preferentially bind to various trans-synaptic partners, including neuroligins, cerebellins, neurexophilins and LRRTMs, each of which subserves different synaptic functions (de Wit and Ghosh, 2016). *NRXN1-α* trans-synaptic interactions play a critical role in thalamo-cortical synaptogenesis and plasticity (Singh et al., 2016), suggesting one possible mechanism in support of the prevailing theory that TS arises from abnormal sensorimotor CSTC circuit development (Jahanshahi et al., 2015).

Although previous studies have observed heterozygous exonic *NRXN1* deletions in TS patients (Fernandez et al., 2012; Nag et al., 2013; Sundaram et al., 2010), small sample sizes precluded a

definitive association of this deletion with TS. We demonstrate, in a large independent sample, that exonic deletions affecting *NRXN1* confer a substantial increase in TS risk. The association of heterozygous *NRXN1* deletions with different NDDs is one of the most reliable findings in the neuropsychiatry CNV literature (Lowther et al., 2017). Consistent with this, 4 of the 12 TS cases with exonic *NRXN1* deletions in our sample had another broadly-defined NDD (2 ASD, 1 DD, 1 Developmental Speech/Language Disorder unspecified) (Table S5), supporting the hypothesis that these deletions may interfere with a generalized neurodevelopmental process which, when combined with other disease-specific mutations and/or background polygenic risk, results in the observed phenotypic pleiotropy.

Like *NRXN1*, *CNTN6* encodes a cell-adhesion molecule expressed primarily in the central nervous system (Ogawa et al., 1996). Contactins are members of the L1 immunoglobulin (Ig) superfamily of proteins, and *Cntn6* has multiple functions in the developing mouse nervous system, including orientation of apical dendrites in cortical pyramidal neurons (Ye et al., 2008), regulation of Purkinje cell development and synaptogenesis (Sakurai et al., 2009), and oligodendrocyte differentiation from neuroprogenitor cells (Cui et al., 2004). Mice with homozygous inactivation of *Cntn6* display delayed corticospinal tract formation and motor impairment (Huang et al., 2012).

Duplications in *CNTN6* represent a novel association for TS. CNVs affecting *CNTN6* have been reported in isolated cases of intellectual disability/developmental delay (ID/DD) (Kashevarova et al., 2014), and deletions alone are enriched in ASD (Mercati et al., 2016). Notably, in a clinical series of 3,724 patients referred for cytogenetic testing, all 7 *CNTN6* duplication carriers either presented with or had a first-degree relative with ADHD and/or OCD, while none of the 7 *CNTN6* deletion carriers were diagnosed with these two common TS comorbidities (Hu et al., 2015). In our study, the rates of co-

morbid OCD/ADHD were not increased in TS *CNTN6* CNV carriers compared to non-carriers, and no TS *CNTN6* carrier was noted to have ASD/ID/DD (Table S5).

There are several limitations of the current study that can inform future inquiry. First, although our sample represents the largest survey of CNVs in TS to date (2,434 cases and 4,093 controls), it is still underpowered to detect extremely rare CNVs and/or those of moderate effect size. Although we show strong evidence for the involvement of deletions across *NRXN1*, our data does not support other previously implicated loci, including deletions in *COL8A1 (Nag et al., 2013)*, observed only once in a single TS patient in this study. While a nominal enrichment of *COL8A1* deletions in TS was originally described in a South American population isolate and possibly represents a population-specific TS risk factor, we emphasize the need for further increases in sample size for continued discovery and refinement of candidate TS loci. Second, while our TS cases were well characterized for OCD and ADHD, we did not formally assess ASD, ID, SCZ or epilepsy. Parents and/or adult subjects were queried about existing diagnoses of these NDDs as well as learning disorders/developmental delay, but cases with milder ASD/DD may not have been detected. Additional efforts should focus on characterizing the full scope of phenotypes associated with *NRXN1* and *CNTN6* CNVs. A comprehensive molecular analysis of these CNVs, including the precise delineation of CNV breakpoints using an auxiliary technology and evaluation of their impact on gene function, will also be needed to understand how these variants increase risk for such phenotype(s). Finally, the elevated burden observed here was largely confined to large singletons and known pathogenic CNVs, consistent with a global enrichment of CNVs under strong negative selection that likely arose *de novo* or within the last few generations. This suggests that, in addition to substantial increases in sample size, alternative study designs that allow for the discrimination of *de novo* CNVs will be fruitful in TS, as has recently been shown for likely-gene-disrupting variants identified by exome sequencing in TS trios (WIllsey et al., 2017).

AUTHOR CONTRIBUTIONS

All authors were involved in the conception and design of this study. A.H., P.P., C.A.M., J.M.S., and G.C. designed and oversaw the analyses. A.H., D.Y., L.K.D., J.H.S., F.T., V.R., I.Z., E.M.R., L.O., J.A.C., L.M.M., B.M.N., N.B.F., P.P., C.A.M., J.M.S., and G.C. conducted the analyses. Major contributions to writing and editing were made by A.H., C.A.M., J.M.S, and G.C. All authors assisted with critically revising the manuscript.

ACKNOWLEDGMENTS

The authors thank the patients with Tourette Syndrome and their families, and all the volunteers who participated in this study. This study was supported by the US NIH U01 NS040024 to Drs. Pauls, Mathews, and Scharf and the TSAICG, ARRA Grant NS040024-09S1, K23 MH085057, and K02 NS085048 to Dr. Scharf, ARRA Grants NS040024-07S1 and NS016648 to Dr. Pauls, MH096767 to Dr. Mathews and NINDS Informatics Center for Neurogenetics and Neurogenomics grant P30 NS062691 to Drs. Coppola and Freimer, by grants from the Tourette Association of America to Drs. Paschou, Pauls, Mathews, and Scharf and from the German Research Society to Dr. Hebebrand.

WORKS CITED

Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.-R., Gunel, M., et al. (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science *310*, 317–320.

Browne, H.A., Hansen, S.N., Buxbaum, J.D., Gair, S.L., Nissen, J.B., Nikolajsen, K.H., Schendel, D.E., Reichenberg, A., Parner, E.T., and Grice, D.E. (2015). Familial clustering of tic disorders and obsessive-compulsive disorder. JAMA Psychiatry *72*, 359–366.

Browne, H.A., Modabbernia, A., Buxbaum, J.D., Hansen, S.N., Schendel, D.E., Parner, E.T., Reichenberg, A., and Grice, D.E. (2016). Prenatal Maternal Smoking and Increased Risk for Tourette Syndrome and Chronic Tic Disorders. J. Am. Acad. Child Adolesc. Psychiatry *55*, 784–791.

Burd, L., Li, Q., Kerbeshian, J., Klug, M.G., and Freeman, R.D. (2009). Tourette syndrome and comorbid pervasive developmental disorders. J. Child Neurol. *24*, 170–175.

Ching, M.S.L., Shen, Y., Tan, W.-H., Jeste, S.S., Morrow, E.M., Chen, X., Mukaddes, N.M., Yoo, S.-Y., Hanson, E., Hundley, R., et al. (2010). Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. Am. J. Med. Genet. B Neuropsychiatr. Genet. *153B*, 937–947.

CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium, and Psychosis Endophenotypes International Consortium (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. *49*, 27–35.

Cui, X.-Y., Hu, Q.-D., Tekaya, M., Shimoda, Y., Ang, B.-T., Nie, D.-Y., Sun, L., Hu, W.-P., Karsak, M., Duka, T., et al. (2004). NB-3/Notch1 pathway via Deltex1 promotes neural progenitor cell differentiation into oligodendrocytes. J. Biol. Chem. *279*, 25858–25865.

Davis, L.K., Yu, D., Keenan, C.L., Gamazon, E.R., Konkashbaev, A.I., Derks, E.M., Neale, B.M., Yang, J., Lee, S.H., Evans, P., et al. (2013). Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. PLoS Genet. *9*, e1003864.

Draper, A., Stephenson, M.C., Jackson, G.M., Pépés, S., Morgan, P.S., Morris, P.G., and Jackson, S.R. (2014). Increased GABA contributes to enhanced control over motor excitability in Tourette syndrome. Curr. Biol. *24*, 2343–2347.

Fernandez, T.V., Sanders, S.J., Yurkiewicz, I.R., Ercan-Sencicek, A.G., Kim, Y.-S., Fishman, D.O., Raubeson, M.J., Song, Y., Yasuno, K., Ho, W.S.C., et al. (2012). Rare copy number variants in tourette syndrome disrupt genes in histaminergic pathways and overlap with autism. Biol. Psychiatry *71*, 392–402.

Fuccillo, M.V., Földy, C., Gökce, Ö., Rothwell, P.E., Sun, G.L., Malenka, R.C., and Südhof, T.C. (2015). Single-Cell mRNA Profiling Reveals Cell-Type-Specific Expression of Neurexin Isoforms. Neuron *87*, 326–340.

Gilbert, D.L., Bansal, A.S., Sethuraman, G., Sallee, F.R., Zhang, J., Lipps, T., and Wassermann, E.M. (2004). Association of cortical disinhibition with tic, ADHD, and OCD severity in Tourette syndrome. Mov.

Disord. *19*, 416–425.

Greene, D.J., Williams, A.C., Iii, Koller, J.M., Schlaggar, B.L., and Black, K.J. (2016). Brain structure in pediatric Tourette syndrome. Mol. Psychiatry.

Hirschtritt, M.E., Lee, P.C., Pauls, D.L., Dion, Y., Grados, M.A., Illmann, C., King, R.A., Sandor, P., McMahon, W.M., Lyon, G.J., et al. (2015). Lifetime prevalence, age of risk, and genetic relationships of comorbid psychiatric disorders in Tourette syndrome. JAMA Psychiatry *72*, 325–333.

Hu, J., Liao, J., Sathanoori, M., Kochmar, S., Sebastian, J., Yatsenko, S.A., and Surti, U. (2015). CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. J. Neurodev. Disord. *7*, 26.

Huang, Z., Yu, Y., Shimoda, Y., Watanabe, K., and Liu, Y. (2012). Loss of neural recognition molecule NB-3 delays the normal projection and terminal branching of developing corticospinal tract axons in the mouse. J. Comp. Neurol. *520*, 1227–1245.

Jahanshahi, M., Obeso, I., Rothwell, J.C., and Obeso, J.A. (2015). A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. Nat. Rev. Neurosci. *16*, 719–732.

Kashevarova, A.A., Nazarenko, L.P., Schultz-Pedersen, S., Skryabin, N.A., Salyukova, O.A., Chechetkina, N.N., Tolmacheva, E.N., Rudko, A.A., Magini, P., Graziano, C., et al. (2014). Single gene microdeletions and microduplication of 3p26.3 in three unrelated families: CNTN6 as a new candidate gene for intellectual disability. Mol. Cytogenet. *7*, 97.

Kearney, H.M., Thorland, E.C., Brown, K.K., Quintero-Rivera, F., South, S.T., and Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee (2011). American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet. Med. *13*, 680–685.

Leivonen, S., Voutilainen, A., Chudal, R., Suominen, A., Gissler, M., and Sourander, A. (2016). Obstetric and Neonatal Adversities, Parity, and Tourette Syndrome: A Nationwide Registry. J. Pediatr. *171*, 213–219.

Lowther, C., Speevak, M., Armour, C.M., Goh, E.S., Graham, G.E., Li, C., Zeesman, S., Nowaczyk, M.J.M., Schultz, L.-A., Morra, A., et al. (2017). Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression. Genet. Med. *19*, 53–61.

Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell *148*, 1223–1241.

Marsh, R., Maia, T.V., and Peterson, B.S. (2009). Functional disturbances within frontostriatal circuits across multiple childhood psychopathologies. Am. J. Psychiatry *166*, 664–674.

Mataix-Cols, D., Isomura, K., Pérez-Vigil, A., Chang, Z., Rück, C., Larsson, K.J., Leckman, J.F., Serlachius, E., Larsson, H., and Lichtenstein, P. (2015). Familial Risks of Tourette Syndrome and Chronic Tic Disorders. A Population-Based Cohort Study. JAMA Psychiatry *72*, 787–793.

McGrath, L.M., Yu, D., Marshall, C., Davis, L.K., Thiruvahindrapuram, B., Li, B., Cappi, C., Gerber, G., Wolf, A., Schroeder, F.A., et al. (2014). Copy number variation in obsessive-compulsive disorder and tourette syndrome: a cross-disorder study. J. Am. Acad. Child Adolesc. Psychiatry *53*, 910–919.

Mercati, O., Huguet, G., Danckaert, A., André-Leroux, G., Maruani, A., Bellinzoni, M., Rolland, T., Gouder, L., Mathieu, A., Buratti, J., et al. (2016). CNTN6 mutations are risk factors for abnormal auditory sensory perception in autism spectrum disorders. Mol. Psychiatry.

Nag, A., Bochukova, E.G., Kremeyer, B., Campbell, D.D., Muller, H., Valencia-Duarte, A.V., Cardona, J., Rivas, I.C., Mesa, S.C., Cuartas, M., et al. (2013). CNV analysis in Tourette syndrome implicates large genomic rearrangements in COL8A1 and NRXN1. PLoS One *8*, e59061.

Ogawa, J., Kaneko, H., Masuda, T., Nagata, S., Hosoya, H., and Watanabe, K. (1996). Novel neural adhesion molecules in the Contactin/F3 subgroup of the immunoglobulin superfamily: isolation and characterization of cDNAs from rat brain. Neurosci. Lett. *218*, 173–176.

Pak, C., Danko, T., Zhang, Y., Aoto, J., Anderson, G., Maxeiner, S., Yi, F., Wernig, M., and Südhof, T.C. (2015). Human Neuropsychiatric Disease Modeling using Conditional Deletion Reveals Synaptic Transmission Defects Caused by Heterozygous Mutations in NRXN1. Cell Stem Cell *17*, 316–328.

Robertson, M.M., Eapen, V., Singer, H.S., Martino, D., Scharf, J.M., Paschou, P., Roessner, V., Woods, D.W., Hariz, M., Mathews, C.A., et al. (2017). Gilles de la Tourette syndrome. Nat Rev Dis Primers *3*, 16097.

Sakurai, K., Toyoshima, M., Ueda, H., Matsubara, K., Takeda, Y., Karagogeos, D., Shimoda, Y., and Watanabe, K. (2009). Contribution of the neural cell recognition molecule NB-3 to synapse formation between parallel fibers and Purkinje cells in mouse. Dev. Neurobiol. *69*, 811–824.

Scharf, J.M., Yu, D., Mathews, C.A., Neale, B.M., Stewart, S.E., Fagerness, J.A., Evans, P., Gamazon, E., Edlund, C.K., Service, S.K., et al. (2013). Genome-wide association study of Tourette's syndrome. Mol. Psychiatry *18*, 721–728.

Scharf, J.M., Miller, L.L., Gauvin, C.A., Alabiso, J., Mathews, C.A., and Ben-Shlomo, Y. (2015). Population prevalence of Tourette syndrome: a systematic review and meta-analysis. Mov. Disord. *30*, 221–228.

Schreiner, D., Nguyen, T.-M., Russo, G., Heber, S., Patrignani, A., Ahrné, E., and Scheiffele, P. (2014). Targeted combinatorial alternative splicing generates brain region-specific repertoires of neurexins. Neuron *84*, 386–398.

Singh, S.K., Stogsdill, J.A., Pulimood, N.S., Dingsdale, H., Kim, Y.H., Pilaz, L.-J., Kim, I.H., Manhaes, A.C., Rodrigues, W.S., Jr, Pamukcu, A., et al. (2016). Astrocytes Assemble Thalamocortical Synapses by Bridging NRX1α and NL1 via Hevin. Cell *164*, 183–196.

Sundaram, S.K., Huq, A.M., Wilson, B.J., and Chugani, H.T. (2010). Tourette syndrome is associated with recurrent exonic copy number variants. Neurology *74*, 1583–1590.

Ushkaryov, Y.A., Petrenko, A.G., Geppert, M., and Südhof, T.C. (1992). Neurexins: synaptic cell surface proteins related to the alpha-latrotoxin receptor and laminin. Science *257*, 50–56.

Verkerk, A.J.M.H., Mathews, C.A., Joosse, M., Eussen, B.H.J., Heutink, P., Oostra, B.A., and Tourette Syndrome Association International Consortium for Genetics (2003). CNTNAP2 is disrupted in a family with Gilles de la Tourette syndrome and obsessive compulsive disorder. Genomics *82*, 1–9.

Westfall, P.H., and Troendle, J.F. (2008). Multiple testing with minimal assumptions. Biom. J. *50*, 745–755.

de Wit, J., and Ghosh, A. (2016). Specification of synaptic connectivity by cell surface interactions. Nat. Rev. Neurosci. *17*, 22–35.

Ye, H., Tan, Y.L.J., Ponniah, S., Takeda, Y., Wang, S.-Q., Schachner, M., Watanabe, K., Pallen, C.J., and Xiao, Z.-C. (2008). Neural recognition molecules CHL1 and NB-3 regulate apical dendrite orientation in the neocortex via PTP alpha. EMBO J. *27*, 188–200.

Willsey, J., et al., 2017, Neuron 94, 1–14 May 3, 2017 http://dx.doi.org/10.1016/j.neuron.2017.04.024

FIGURE LEGENDS

Figure 1. Flow chart of experimental procedures and analyses. CNVs were called from genome-wide SNP genotype data generated from 2,434 TS cases and 4,093 controls (grey). Data processing, CNV detection and quality control steps (blue) are described in the STAR Methods. An outline of the main analyses are presented in red. Figures or tables relevant to each outlined step are in parentheses.

Figure 2. Rare CNV burden in 2,434 TS cases and 4,093 controls.

(A) The global burden of all rare (<1% frequency) CNVs > 30kb is shown for genic (top) and non-genic (bottom) CNVs and stratified by CNV type (all, loss (deletions), gain (duplications)). Global CNV burden is compared using three different metrics: 1) CNV Count, total number of CNVs per subject; 2) CNV length, aggregate length of all CNVs (in Mb); 3) CNV gene count, number of genes spanned by CNVs. Control rate, averaged baseline burden metric per control subject. Red boxes, odds ratios (box size is proportional to standard error); Blue lines, 95% confidence intervals. Genic CNVs are defined as those that overlap any exon of a known protein-coding gene.

(B) Analyses in (A) were assessed further by partitioning CNV length burden of all CNVs (deletions + duplications) into different CNV size categories. Whiskers represent 95% confidence intervals.

(C) The analysis in (B) was repeated for CNVs binned by frequency.

Odds ratios (OR) were calculated from logistic regression analyses adjusted for covariates (see STAR Methods) using standardized burden metrics to allow for comparison. ORs >1 indicate an increased TS risk. P-values are derived from the likelihood ratio test.

Figure 3. Large, singleton CNVs and known pathogenic variants are overrepresented in TS.

(A) CNV count burden restricted to singleton events, stratified by CNV size and type (deletion/duplication). (B) CNV burden of all rare CNVs, separated by clinical relevance (benign, uncertain, pathogenic) according to the American College of Medical Genetics guidelines. ORs > 1 represent an increase in risk for TS per CNV. P-values are derived from the likelihood ratio test.

Figure 4. Segmental and gene-based tests converge on two distinct loci significantly enriched in TS cases.

(A) Manhattan plot of segmental association test results representing genome-wide corrected p-values calculated at each CNV breakpoint. The two genome-wide significant association peaks correspond to deletions at *NRXN1* ($P_{locus}=7.0x10^{-6}$, $P_{corr}=1.0x10^{-3}$) and duplications at *CNTN6* ($P_{locus}=5.4x10^{-5}$, $P_{corr}=6.9x10^{-3}$). Red and blue levels correspond to a genome-wide corrected α of 0.05 and 0.01, respectively.

(B) Heterozygous exonic deletions in *NRXN1* found in 12 cases (0.49%) and 1 control (0.03%), corresponding to an OR=20.2, 95% CI (2.6-155.2). Exon-affecting CNVs cluster at the 5' end with deletions across exons 1-3 found in 10 cases and no controls. Red, deletions in TS cases; Dark red, deletion in controls; Blue, case duplication.

(C) Exon-spanning duplications over *CNTN6* found in 12 cases and 2 controls (OR=10.2, [2.0-17.8]) *CNTN6* duplications are considerably larger in cases compared to controls (640 vs. 143 kb, on average). Blue, case duplications; Dark blue, control duplications; Red, case deletion; Dark Red, control duplication.

Table S1. Sample genotyping and QC summary

Table S1A: Studies and genotyping			
GENOTYPING BATCH	**ARRAY**	**CENTER**	**PHENOTYPES**
CC	OmniExpress v 1.0	Cardiff	Control
CNP	OmniExpress v 1.0	Broad	Control/Clinical
GPC	OmniExpress v 1.0	Broad	Control/Clinical
WTCCC2	OmniExpress v 1.0	Cardiff	Control
GGRI1	OmniExpress Exome v 1.1	UCLA	Control/TS
GGRI2	OmniExpress Exome v 1.1	UCLA	Control/TS
GGRI3	OmniExpress Exome v 1.1	UCLA	Control/TS

Table S1B: QC summary								
QC STEP	**CC**	**CNP**	**GPC**	**WTCCC2**	**GGRI1**	**GGRI2**	**GGRI3**	**TOTALS**
Initial Samples	1,146	1,511	3,197	960	1,152	2,160	136	10,262
Pre-cluster QC	1,141	1,510	3,126	870	1,148	2,152	135	10,082
Duplicate/Control Samples	1,141	1,491	3,081	870	1,134	2,143	134	9,994
Clinical Phenotype	1,141	1,312	1,388	870	1,134	2,143	134	8,122
Sex Concordance	1,141	1,312	1,387	870	1,132	2,140	134	8,116
Heterozygosity	1,138	1,247	1,301	859	1,107	2,089	133	7,874
Cryptic Relatedness	1,135	1,222	1,264	852	1,100	2,067	132	7,772
Intensity QC	1,101	1,106	1,164	832	1,014	1,843	119	7,179
EU Ancestry	1,067	634	1,143	813	959	1,803	116	6,535
CNV Load QC	1,067	634	1,141	808	958	1,803	116	6,527

Table S2: GMM-based genotype calls at common CNV loci

CNV_ID	CLUSTER	CLUSTER_LRR	GMM_COPY	CTRL_CALLS	CASE_CALLS	CTRL_FREQ	CASE_FREQ	p-value
HM3_CNP_134	1	-13.17478812	0	7	4	0.002	0.002	1.0
HM3_CNP_134	2	-1.544234008	1	296	191	0.072	0.078	0.4
HM3_CNP_156	1	-1.141264855	1	517	315	0.126	0.129	0.7
HM3_CNP_156	2	-10.96495207	0	18	13	0.004	0.005	0.6
HM3_CNP_299	1	-2.148959932	1	275	142	0.067	0.058	0.2
HM3_CNP_299	2	-20.27406193	0	4	3	0.001	0.001	0.7
HM3_CNP_369	1	2.201328191	3	234	137	0.057	0.056	0.9
HM3_CNP_494	1	-2.7402128	1	196	100	0.048	0.041	0.2
HM3_CNP_494	2	-23.74078464	0	1	1	0.000	0.000	1.0
HM3_CNP_540	1	1.648736036	3	392	263	0.096	0.108	0.1
HM3_CNP_540	2	-3.301784945	1	32	17	0.008	0.007	0.8
HM3_CNP_618	1	-3.470922513	1	44	24	0.011	0.010	0.8
HM3_CNP_618	2	1.817743156	3	167	91	0.041	0.037	0.5
HM3_CNP_655	1	-3.645609914	1	45	34	0.011	0.014	0.3
HM3_CNP_692	0	-4.175170396	1	10	11	0.002	0.005	0.2
HM3_CNP_692	2	2.34262532	3	47	32	0.011	0.013	0.6
HM3_CNP_803	1	-10.64502833	0	15	14	0.004	0.006	0.2
HM3_CNP_803	2	-1.347106673	1	417	258	0.102	0.106	0.6
HM3_CNP_850	1	-31.82000658	0	1	1	0.000	0.000	1.0
HM3_CNP_850	2	-2.649145422	1	74	49	0.018	0.020	0.6
HM3_CNP_850	3	1.410233747	3	165	101	0.040	0.041	0.8

Table S3. Sensitivity analysis of HMM-based segmentation calls.

Table S3A: Sensitivity analysis by locus

CNV_ID	CNV_TYPE	GMM_TOTAL	GMM_CTRL	GMM_CASE	HMM_CTRL	HMM_CASE	CTRL_SENSE	CASE_SENSE	p-value
HM3_CNP_134	DEL	498	303	195	300	194	0.99	0.995	1.0
HM3_CNP_156	DEL	863	535	328	531	321	0.993	0.979	0.11
HM3_CNP_299	DEL	424	279	145	279	145	1.000	1.000	1.0
HM3_CNP_369	DUP	371	234	137	208	122	0.889	0.891	1.0
HM3_CNP_494	DEL	298	197	101	197	101	1.000	1.000	1.0
HM3_CNP_540	DUP	655	392	263	391	261	0.997	0.992	0.57
HM3_CNP_540	DEL	49	32	17	32	17	1.000	1.000	1.0
HM3_CNP_618	DEL	68	44	24	44	24	1.000	1.000	1.0
HM3_CNP_618	DUP	258	167	91	166	90	0.994	0.989	1.0
HM3_CNP_655	DEL	79	45	34	45	34	1.000	1.000	1.0
HM3_CNP_692	DEL	21	10	11	10	11	1.000	1.000	1.0
HM3_CNP_692	DUP	79	47	32	47	32	1.000	1.000	1.0
HM3_CNP_803	DEL	704	432	272	428	272	0.991	1.000	0.16
HM3_CNP_850	DEL	125	75	50	75	50	1.000	1.000	1.0
HM3_CNP_850	DUP	266	165	101	164	98	0.994	0.970	0.15

Table S3B: Overall sensitivity across common CNVs

CNV_TYPE	GMM_TOTALS	GMM_CTRL	GMM_CASE	HMM_CTRL	HMM_CASE	CTRL_SENSE	CASE_SENSE	p-value
DEL+DUP	4758	2957	1801	2917	1772	0.986	0.984	0.53
DEL	3129	1952	1177	1941	1169	0.994	0.993	0.81
DUP	1629	1005	624	976	603	0.971	0.966	0.65

Table S3C: Group-wise sensitivity analysis across individuals

CNV_TYPE	CTRL_SENSE	Std. Error	CASE_SENSE	Std. Error	p-value
DEL+DUP	0.989	0.002	0.983	0.003	0.15
DEL	0.996	0.001	0.991	0.002	0.14
DUP	0.973	0.005	0.967	0.007	0.46

Table S5: Clinical phenotypes of *NRXN1* and *CNTN6* CNV carriers

Sample ID	Gene	Chr	Start	End	Type	Length	Variant Effect	OCD	ADHD	Atypical	Notes
TS1_0630	NRXN1	2	50821559	51021488	DEL	199.9	CODING	N	N	Y	Unspecified Developmental Delay (ICD-9: 315.9)
TS1_0180	NRXN1	2	50930181	51272375	DEL	342.2	CODING	N	N	Y	Asperger Syndrome
TS1_0446	NRXN1	2	50945471	51770480	DEL	825	CODING	Y	Y	N	
TS1_0105	NRXN1	2	51002606	51316822	DEL	314.2	CODING	N	Y	N	
TS2_1256	NRXN1	2	51028662	51458570	DEL	429.9	CODING	N	Y	Y	Other developmental speech or language disorder (ICD-9: 315.39)
TS2_0026	NRXN1	2	51041472	51483528	DEL	442.1	CODING	N	N	N	
TS2_0924	NRXN1	2	51041603	51528298	DEL	486.7	CODING	N	Y	N	
TS2_0750	NRXN1	2	51058745	51252137	DEL	193.4	CODING	Y	Y	Y	Asperger Syndrome
TS2_1238	NRXN1	2	51077569	51458570	DEL	381	CODING	Y	N	Y	Paranoid personality disorder
TS1_0573	NRXN1	2	51079482	51357902	DEL	278.4	CODING	NA	NA	NA	
TS1_0776	NRXN1	2	51101583	51308895	DEL	207.3	CODING	N	Y	N	Brother with Asperger Syndrome
TS1_0698	NRXN1	2	51123048	51286169	DEL	163.1	CODING	Y	Y	N	
TS2_1805	CNTN6	3	565961	1350458	DUP	784.5	CODING	Y	NA	N	
TS2_1405	CNTN6	3	668832	1143424	DUP	474.6	5' UTR	Y	Y	N	
TS2_1624	CNTN6	3	707257	1781739	DUP	1074	CODING	N	N	N	
TS2_1525	CNTN6	3	857325	1427769	DUP	570.4	CODING	Y	Y	N	
TS2_1568	CNTN6	3	864513	1425997	DUP	561.5	CODING	Y	Y	N	
TS2_1545	CNTN6	3	864513	1427769	DUP	563.3	CODING	N	Y	N	
TS2_1320	CNTN6	3	946290	1276092	DUP	329.8	CODING	Y	N	N	
TS1_0618	CNTN6	3	1125605	1315900	DUP	190.3	CODING	N	N	N	
TS2_1156	CNTN6	3	1218279	2170519	DUP	952.2	CODING	N	Y	N	
TS1_0558	CNTN6	3	1218279	2170519	DUP	952.2	CODING	N	Y	N	
TS2_0827	CNTN6	3	1226953	2170519	DUP	943.6	CODING	N	N	N	
TS2_0452	CNTN6	3	1260932	1556680	DUP	295.7	CODING	N	N	N	

Figure 1

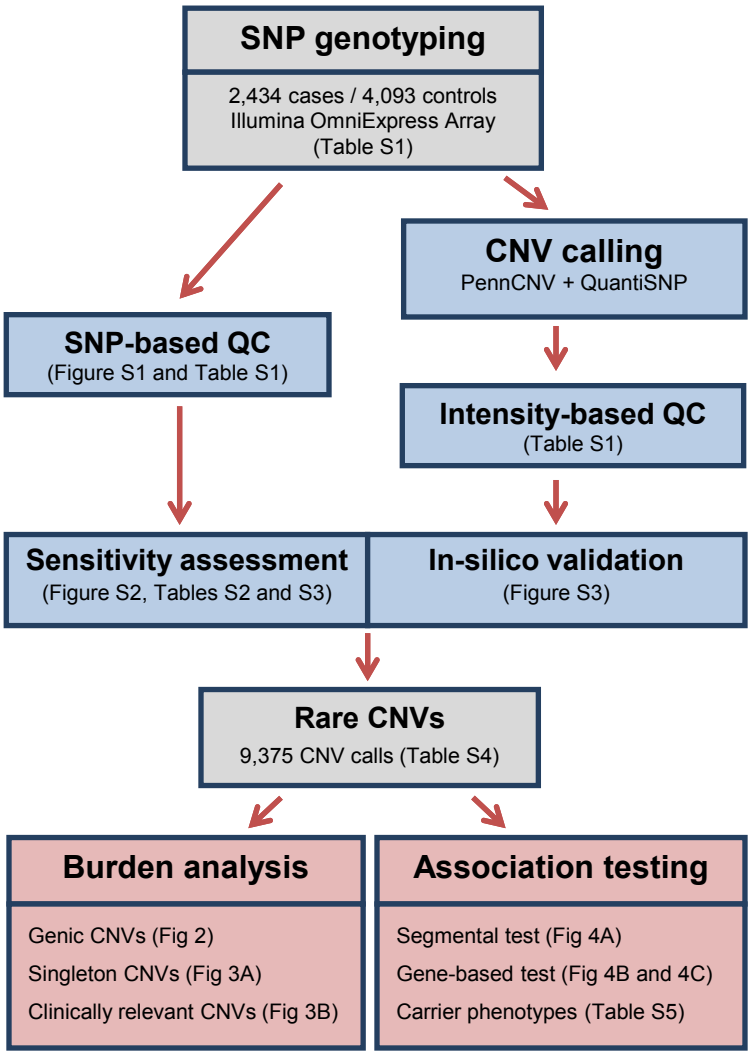
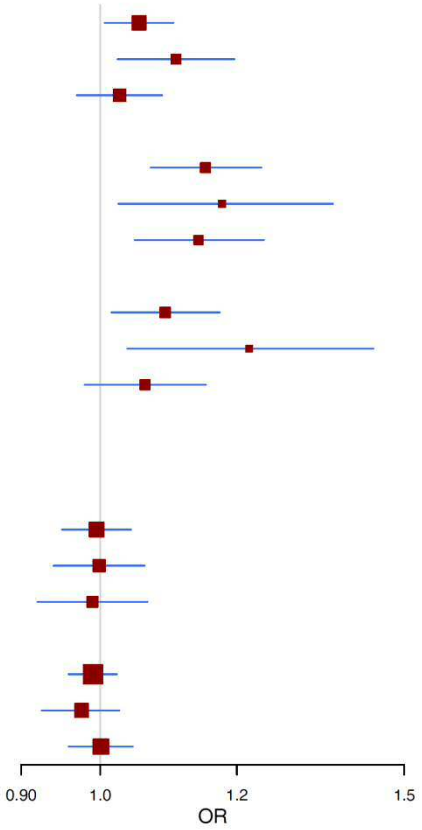

SNP genotyping
2,434 cases / 4,093 controls
Illumina OmniExpress Array
(Table S1)
CNV calling
PennCNV + QuantiSNP
SNP-based QC
(Figure S1 and Table S1)
Intensity-based QC
(Table S1)
Sensitivity assessment
(Figure S2, Tables S2 and S3)
In-silico validation
(Figure S3)
Rare CNVs
9,375 CNV calls (Table S4)
Burden analysis
Genic CNVs (Fig 2)
Singleton CNVs (Fig 3A)
Clinically relevant CNVs (Fig 3B)
Association testing
Segmental test (Fig 4A)
Gene-based test (Fig 4B and 4C)
Carrier phenotypes (Table S5)

Figure 2

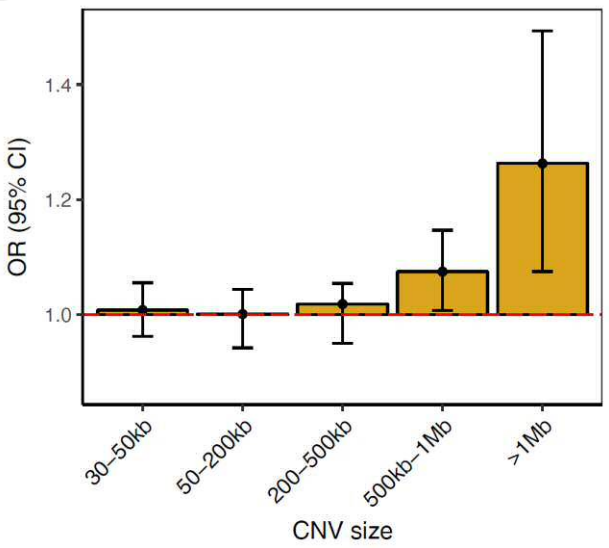
A

Genic CNVs	Control rate	OR	95% CI	p–val
CNV count				
all	0.70	1.05	[1.01–1.10]	0.027
loss	0.24	1.11	[1.02–1.19]	0.011
gain	0.46	1.03	[0.97–1.09]	0.372
CNV length				
all	1.58	1.15	[1.07–1.24]	1.9e–4
loss	0.43	1.18	[1.02–1.36]	0.026
gain	1.15	1.14	[1.05–1.24]	2.9e–3
CNV gene count				
all	1.78	1.09	[1.01–1.17]	0.019
loss	0.44	1.22	[1.04–1.44]	0.017
gain	1.34	1.06	[0.98–1.15]	0.147

Non–genic CNVs	Control rate	OR	95% CI	p–val
CNV count				
all	0.77	1.00	[0.95–1.04]	0.839
loss	0.46	1.00	[0.94–1.06]	0.967
gain	0.31	0.99	[0.92–1.07]	0.786
CNV length				
all	0.97	0.99	[0.96–1.02]	0.564
loss	0.51	0.98	[0.93–1.03]	0.344
gain	0.46	1.00	[0.96–1.04]	0.963



B



C

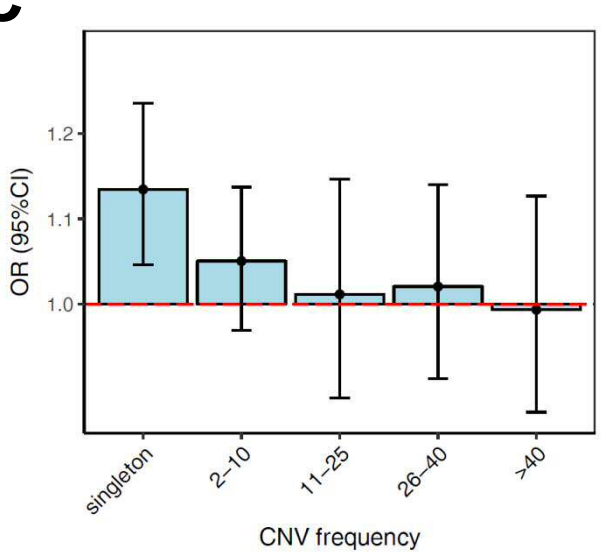
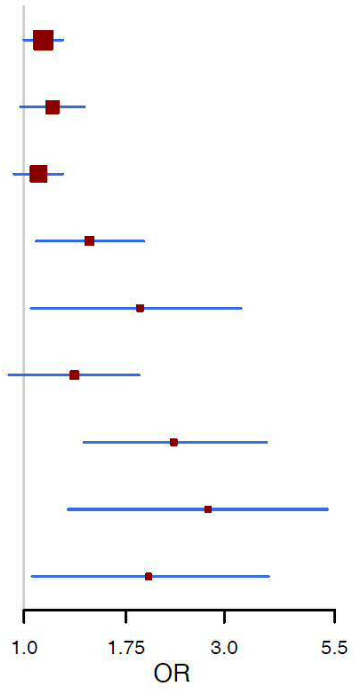


Figure 3

A

Burden of singleton CNVs

Size	Type	Control rate	OR	95% CI	p–val
<500kb	all	0.251	1.11	[1.00–1.23]	0.050
	loss	0.090	1.17	[0.98–1.39]	0.082
	gain	0.161	1.08	[0.95–1.24]	0.243
>500kb	all	0.029	1.43	[1.07–1.93]	0.010
	loss	0.006	1.89	[1.04–3.29]	0.022
	gain	0.023	1.32	[0.92–1.88]	0.129
>1MB	all	0.008	2.28	[1.39–3.79]	1.2e–03
	loss	0.003	2.75	[1.28–5.23]	6.5e–03
	gain	0.005	1.98	[1.04–3.83]	0.038



B

Burden by clinical relevance

ACMG class	Type	Control rate	OR	95% CI	p–val
BENIGN	all	0.101	1.07	[0.91–1.24]	0.417
	loss	0.037	1.02	[0.77–1.33]	0.907
	gain	0.069	1.09	[0.90–1.32]	0.357
UNCERTAIN	all	0.496	1.04	[0.96–1.12]	0.344
	loss	0.150	1.12	[0.97–1.28]	0.122
	gain	0.346	1.00	[0.91–1.10]	0.922
PATHOGENIC	all	0.008	3.03	[1.85–5.07]	1.5e–5
	loss	0.003	3.94	[1.83–8.95]	6.3e–4
	gain	0.005	2.48	[1.31–4.84]	5.9e–3

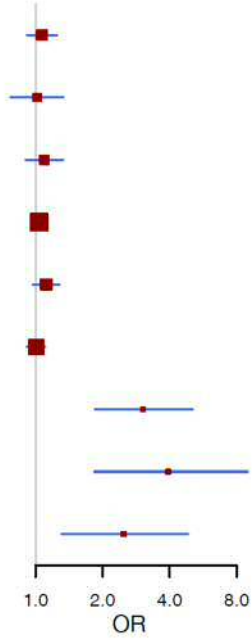
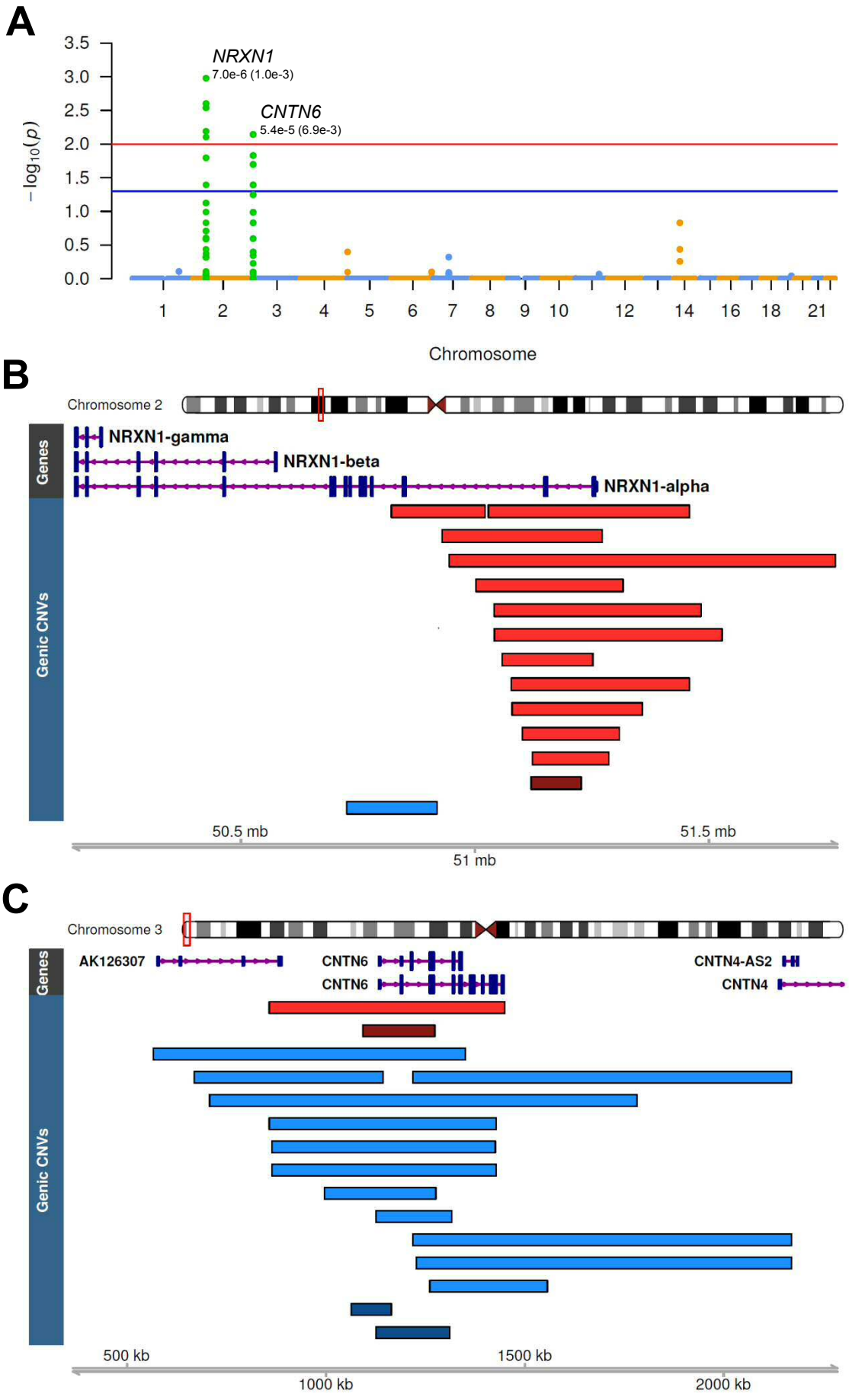


Figure 4


A
NRXN1
7.0e-6 (1.0e-3)
CNTN6
5.4e-5 (6.9e-3)
−log10(p)
Chromosome
B
Chromosome 2
Genes
Genic CNVs
NRXN1-gamma
NRXN1-beta
NRXN1-alpha
50.5 mb
51.5 mb
51 mb
C
Chromosome 3
Genes
Genic CNVs
AK126307
CNTN6
CNTN6
CNTN4-AS2
CNTN4
500 kb
1500 kb
1000 kb
2000 kb

SUPPLEMENTAL INFORMATION

Supplemental Figures

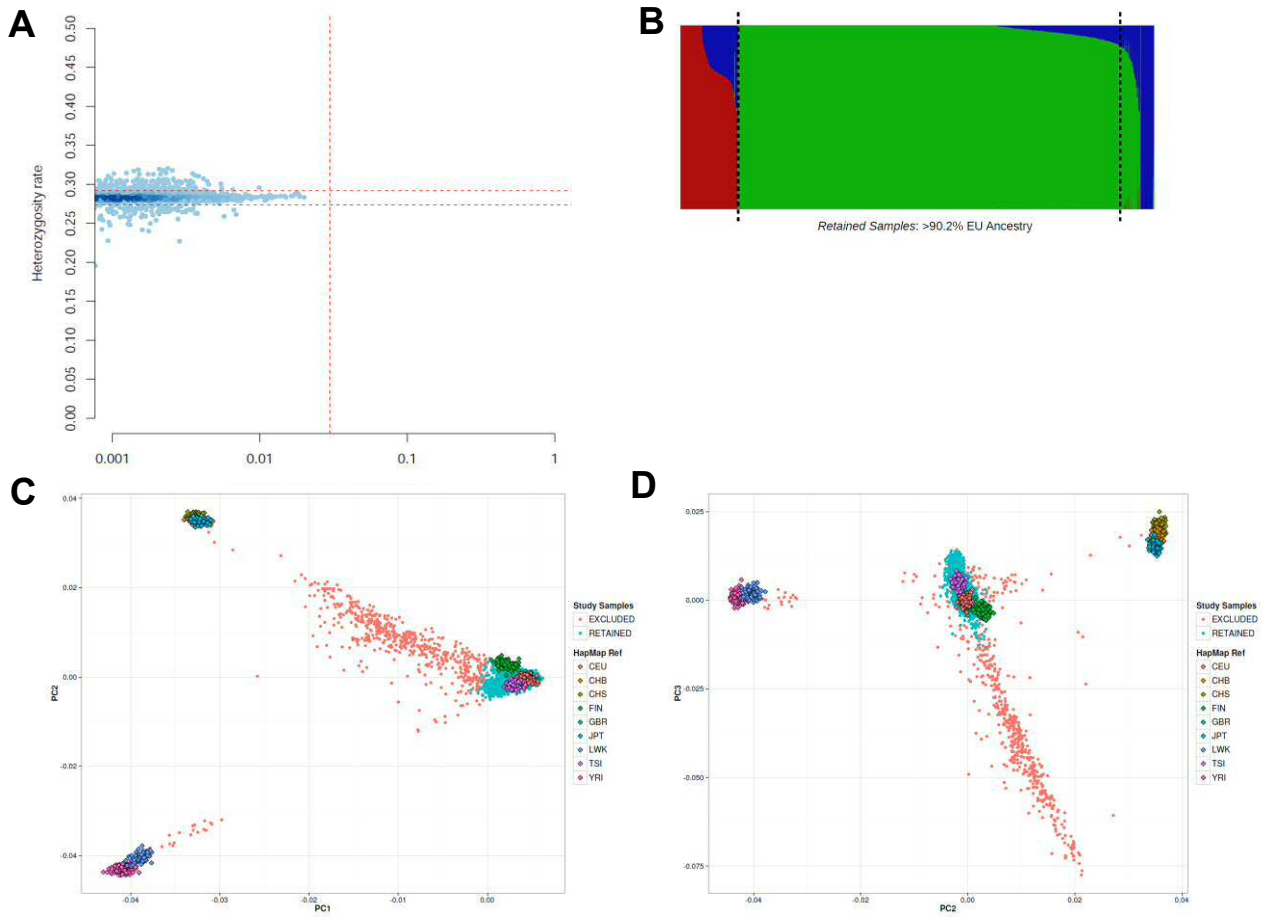


Figure S1. SNP-based quality control and ancestry determination (related to Figure 1). (A) Exclusion of sample outliers based on heterozygosity, mean +/- 1.5 SD (red dotted lines). (B) Exclusion of non-European samples based on ethnicity estimation using fastStructure with HapMap continental groups and K=3 clustering. Samples with > 9.85% non-EU ancestry were excluded. This threshold was calibrated against the maximum of reference HapMap/1000 Genomes European groups CEU, GBR, and TSI. The results of principal component (PC) analysis for the cohort and reference groups are plotted along (C) PCs 1 and 2 and (D) PCs 2 and 3. Retained samples and excluded samples are shown in cyan and pink, respectively. CEU,

Utah residents with Northern and Western European ancestry from the CEPH collection; CHB, Han Chinese in Beijing, China; CHS, Southern Han Chinese; FIN, Finnish in Finland; GBR, British in England and Scotland; JPT, Japanese in Tokyo, Japan; LWK, Luhya in Webuye, Kenya; TSI, Toscani in Italia; YRI, Yoruba in Ibadan, Nigeria.

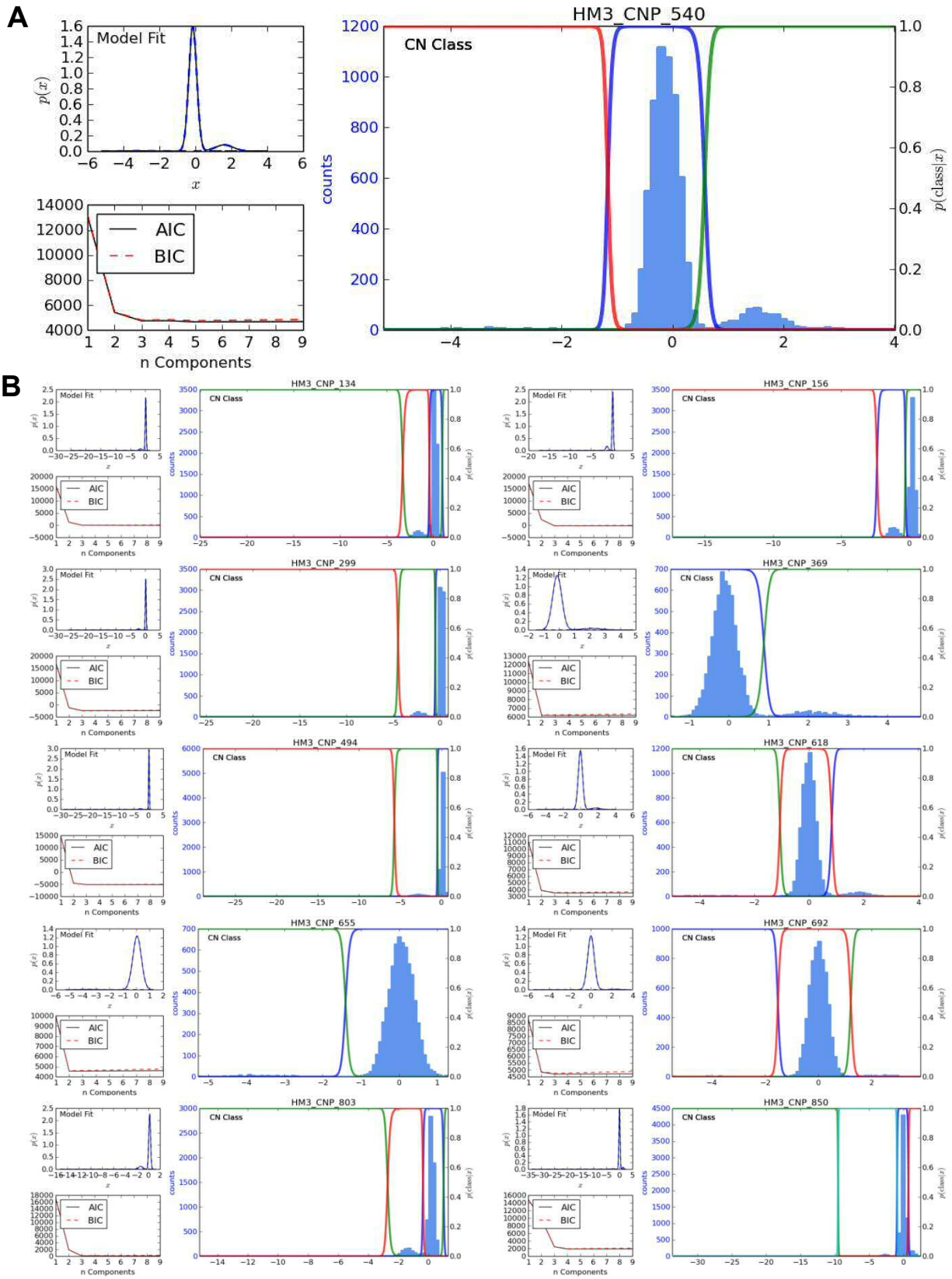
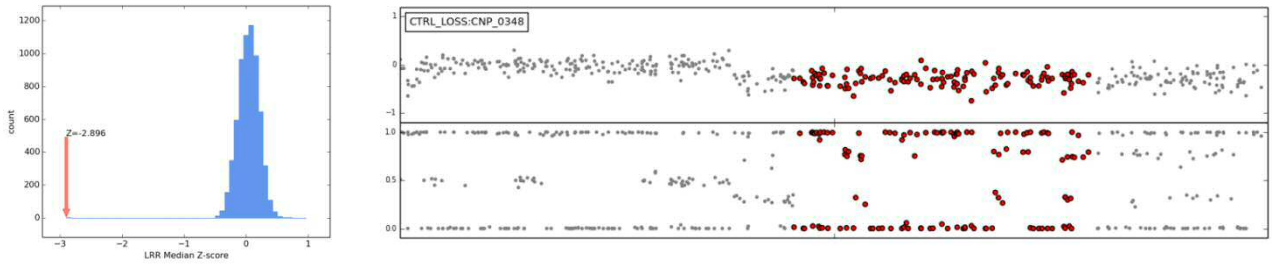

A
Model Fit
p(x)
x
AIC
BIC
n Components
HM3_CNP_540
CN Class
counts
p(class|x)
B
HM3_CNP_134
HM3_CNP_156
HM3_CNP_299
HM3_CNP_369
HM3_CNP_494
HM3_CNP_618
HM3_CNP_655
HM3_CNP_692
HM3_CNP_803
HM3_CNP_850

Figure S2. Gaussian mixture model (GMM) clusters of common HapMap3 CNVs (related to Figure 1, Table S2 and S3). (A) A representative GMM cluster plot for locus HM3_CNP_540. Subplots for each CNV depict, counter-clockwise: the best-fit model, Akaike and Bayesian Information Criterion metrics calculated for GMM fitting 1-9 components, and the posterior probability for CNV cluster assignment (colored lines) overlaying the distribution of median summarized intensity values for all samples across region calculated using the best-fit model. (B) GMM plots for the 10 additional HapMap3 CNV loci that were used to critically evaluate sensitivity between cases and controls (STAR Methods, Table S2 and S3).

A

SAMPLE_ID	CHR	BP1	BP2	COPY	#SNPS	START_SNP	END_SNP	LRR-Z	BAF_{del}	BAF_{dup}	OUTLIER-Z
CNP_0348	6	67801176	67887156	1	21	rs9363696	rs16899159	-2.96	0.62	0.38	0.00015
CNP_0348[1]	**6**	**67907952**	**68586809**	**1**	**120**	**rs12197620**	**rs9354637**	**-2.9**	**0.76**	**0.16**	**0.00015**
CNP_0348	6	68707131	69142008	1	96	rs4707250	rs9363918	-2.93	0.73	0.17	0.00015
WT_0533[2]	**10**	**47375657**	**47703869**	**3**	**48**	**rs28599894**	**rs4434935**	**2.48**	**0.33**	**0.63**	**0.09**
CC_0852	10	47375657	47703869	3	48	rs28599894	rs4434935	2.36	0.33	0.60	0.09
WT_0866	10	47375657	47703869	3	48	rs28599894	rs4434935	2.33	0.38	0.60	0.09
TS_0457	10	47375657	47703869	3	48	rs28599894	rs4434935	2.29	0.39	0.60	0.09
TS_1843	10	47375657	47703869	3	48	rs28599894	rs4434935	2.05	0.39	0.60	0.09

B



C

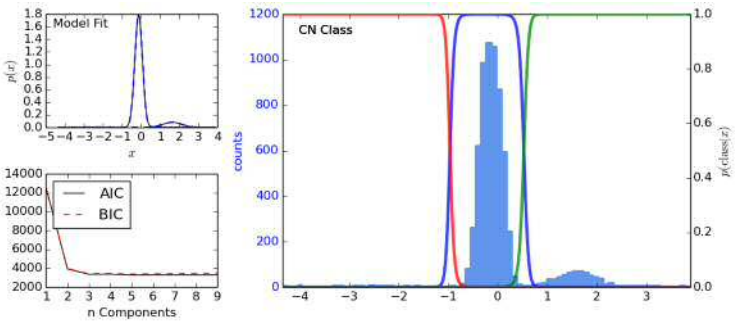


Figure S3. *In silico* validation of CNV calls (related to Figure 1). (A) Representative CNVs scored with various CNV validation metrics. Abbreviations (see STAR Methods for details): median summarized intensity measures across a putative CNV locus, standardized by sample (LRR-Z), proportion of probes with a B Allele Frequency (BAF) banding pattern indicative of a duplication event (BAF-D), proportion of samples with LRR-Z scores indicative of a polymorphic event (OUTLIER-Z). (B) Distribution of median summarized standardized intensity values (LRR-Z) for validated CNVs derived from HapMap samples generated on identical arrays as used in this study (Illumina OmniExpress). Based on this distribution, CNVs without an LRR-Z score <

2.3 (deletions) and > 1.3 (duplications) were flagged for manual inspection. (B) Example of a large singleton mosaic event flagged for exclusion in sample CNP_0348, indicated as (1) in Figure S3A. This CNV on chromosome 6 was detected as three separate CNVs after taking the consensus of two different HMM calling algorithms. The largest CNV call exhibits an LRR-Z score of -2.86 (left, red arrow), indicative of a deletion, but shows a clear BAF-banding pattern of a duplication event (right), with a BAF_{dup} score of 0.16. This is indicative of a mosaic event, where only a proportion of cells from sample CNP_0348 harbor the deletion event. (C) Example of a polymorphic CNV on chr10:47,375,657-47,703,869 misclassified as a rare event due to reduced sensitivity, indicated as (2) in Figure S3A, with an OUTLIER-Z score of 0.09. Genotyping using GMM-based clustering indicated that this misclassified rare event (MAF < 0.01) has a MAF of 0.12.

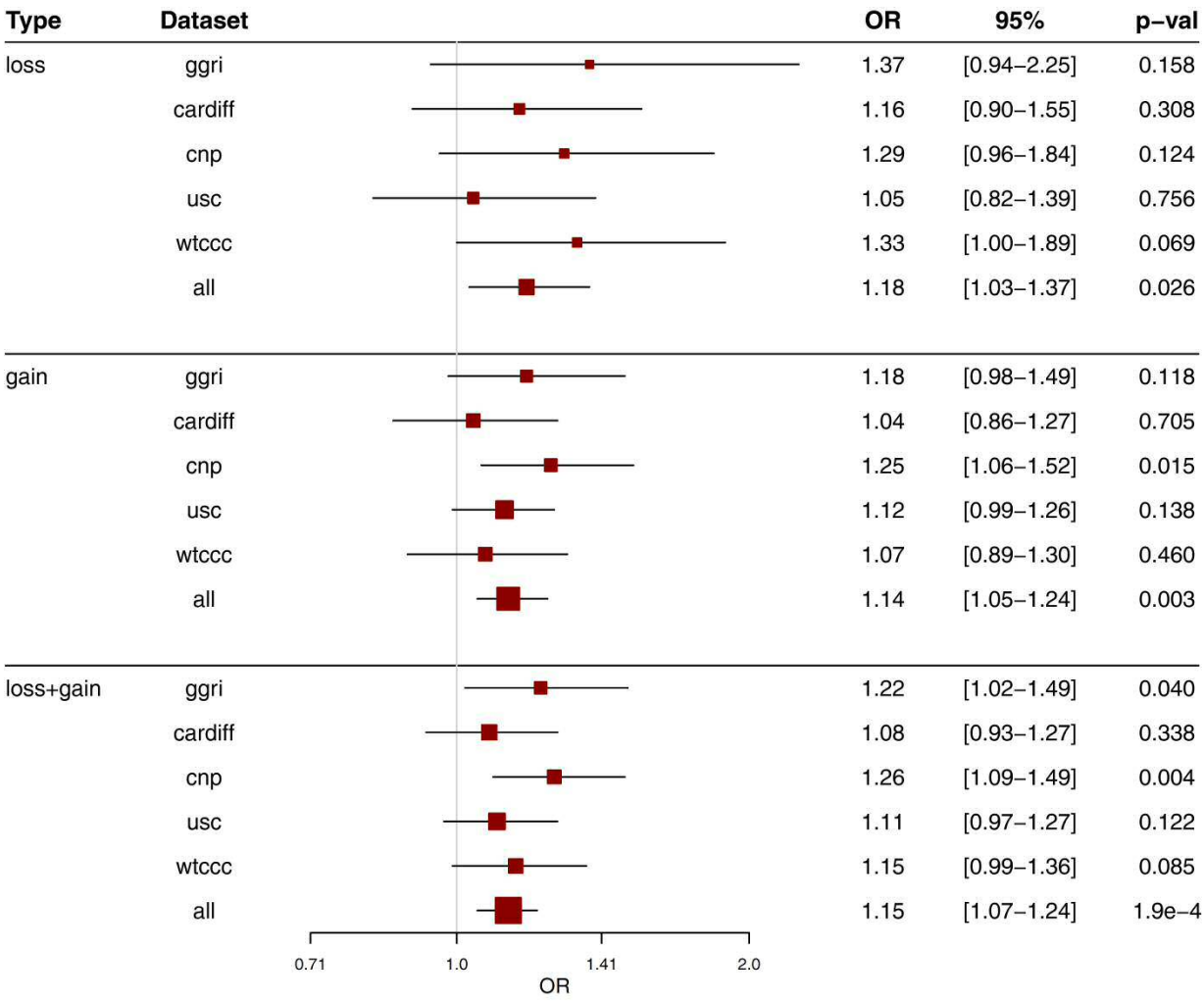


Figure S4. Elevated CNV burden is consistent across datasets (related to Figure 2). We assessed for increased CNV burden using different metrics and found that total CNV length was most significantly associated with an increased risk for TS (Figure 2). To ensure that the enrichment signal was not driven by a single dataset, here we repeated the assessment of burden by total CNV length, examining all TS samples compared to each of the control sample sets individually and to all control samples together. An increased burden is consistent across all datasets, and additionally when stratified by CNV type: loss (deletions); gain (duplications) and loss + gain (both deletions and duplications). GGRI, controls collected and genotyped alongside TS cases; Cardiff, CNP, USC, WTCC, control samples taken from external datasets, as defined in Table S1A and STAR Methods.

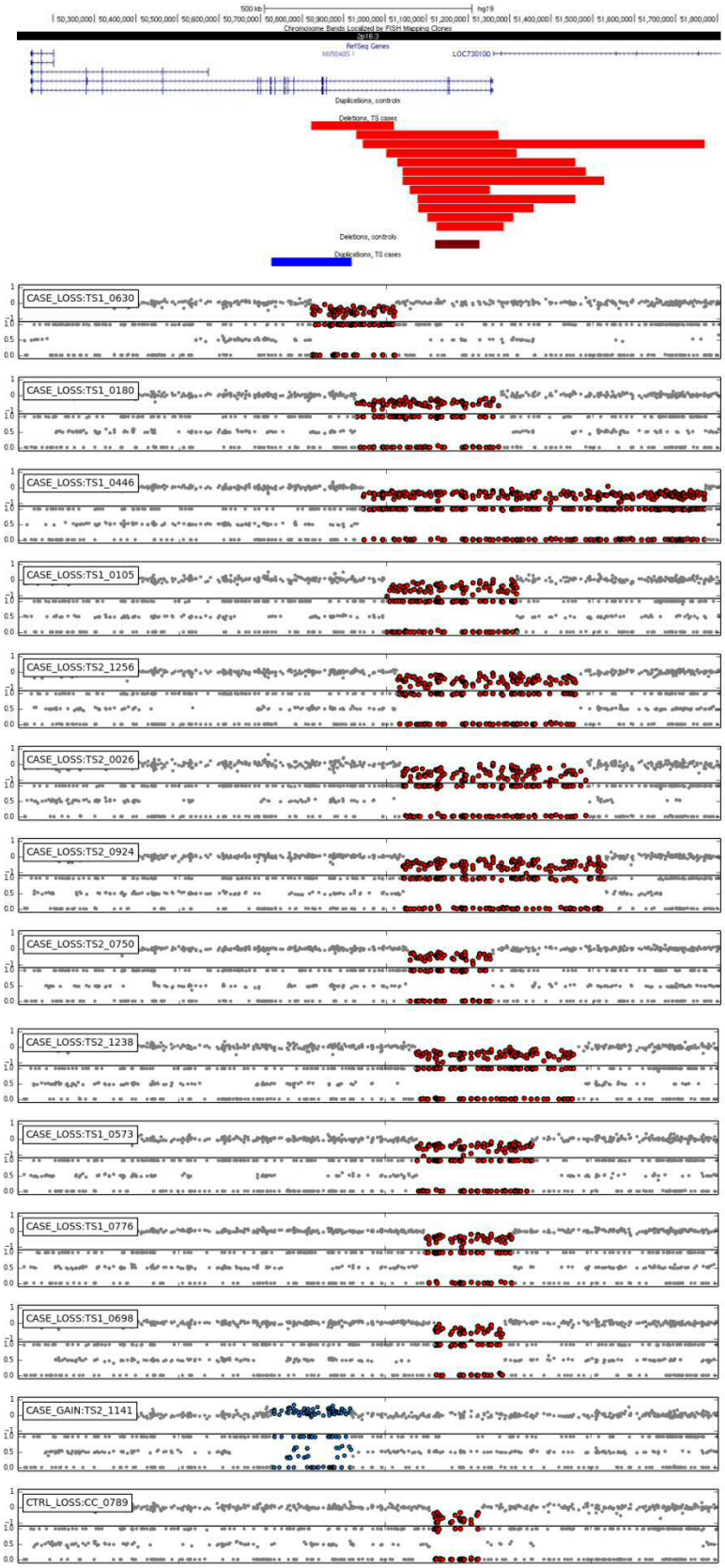

500 kb
hg19
50,300,000 50,400,000 50,500,000 50,600,000 50,700,000 50,800,000 50,900,000 51,000,000 51,100,000 51,200,000 51,300,000 51,400,000 51,500,000 51,600,000 51,700,000 51,800,000
Chromosome Bands Localized by FISH Mapping Clones
20152.3
RefSeq Genes
MIR6485
LOC730100
Duplications, controls
Deletions, TS cases
Deletions, controls
Duplications, TS cases
CASE_LOSS:TS1_0630
CASE_LOSS:TS1_0180
CASE_LOSS:TS1_0446
CASE_LOSS:TS1_0105
CASE_LOSS:TS2_1256
CASE_LOSS:TS2_0026
CASE_LOSS:TS2_0924
CASE_LOSS:TS2_0750
CASE_LOSS:TS2_1238
CASE_LOSS:TS1_0573
CASE_LOSS:TS1_0776
CASE_LOSS:TS1_0698
CASE_GAIN:TS2_1141
CTRL_LOSS:CC_0789

Figure S5. Exonic CNVs affecting *NRXN1* (related to Figure 4). UCSC genome browser track depicting all exonic *NRXN1* CNVs > 50kb identified in this study: 12 heterozygous case deletions (red), one control deletion (dark red) and a single case duplication (blue). Probe-level plots of Log R Ratio (LRR) intensity and B Allele Frequency (BAF) for all exonic *NRXN1* CNV carriers shown in the same order as the UCSC genome browser track. Colored probes indicate the location of called deletions (red) and duplications (blue).

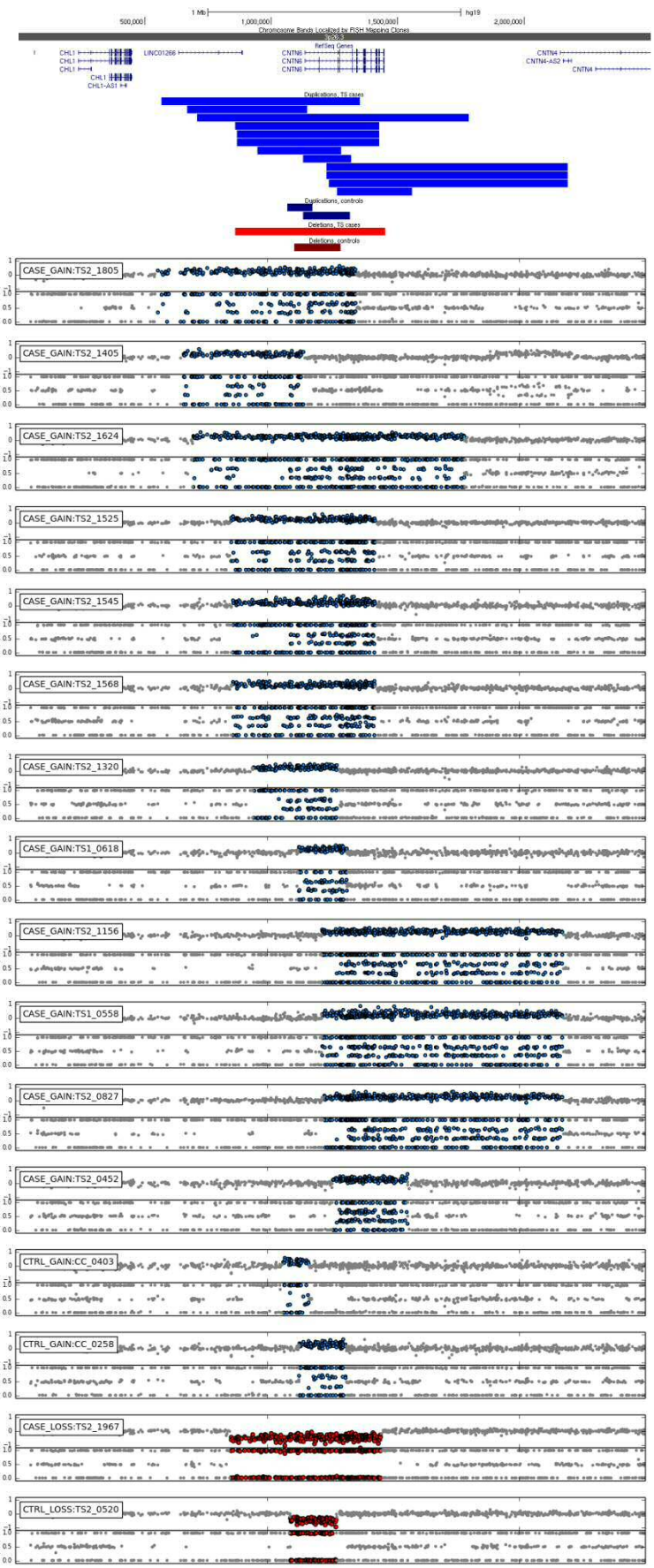

1 Mb
hg19
500,000
1,000,000
1,500,000
2,000,000
Chromosome Bands Localized by FISH Mapping Clones
RefSeq Genes
CHL1
CHL1
CHL1
CHL1
CHL1-AS1
LINC01266
CNTN6
CNTN6
CNTN6
CNTN4
CNTN4-AS2
CNTN4
Duplications, TS cases
Duplications, controls
Deletions, TS cases
Deletions, controls
CASE_GAIN:TS2_1805
CASE_GAIN:TS2_1405
CASE_GAIN:TS2_1624
CASE_GAIN:TS2_1525
CASE_GAIN:TS2_1545
CASE_GAIN:TS2_1568
CASE_GAIN:TS2_1320
CASE_GAIN:TS1_0618
CASE_GAIN:TS2_1156
CASE_GAIN:TS1_0558
CASE_GAIN:TS2_0827
CASE_GAIN:TS2_0452
CTRL_GAIN:CC_0403
CTRL_GAIN:CC_0258
CASE_LOSS:TS2_1967
CTRL_LOSS:TS2_0520

Figure S6. CNVs overlapping *CNTN6* (related to Figure 4). UCSC genome browser track displaying heterozygous genic duplications in TS cases (blue) and controls (dark blue) followed by deletions (red). Probe-level LRR and BAF plots for all 16 CNVs detected spanning *CNTN6* are shown below the genome browser schematic. CNV carriers are shown in the same order as the UCSC genome browser track. Colored probes indicate the location of called deletions (red) and duplications (blue).

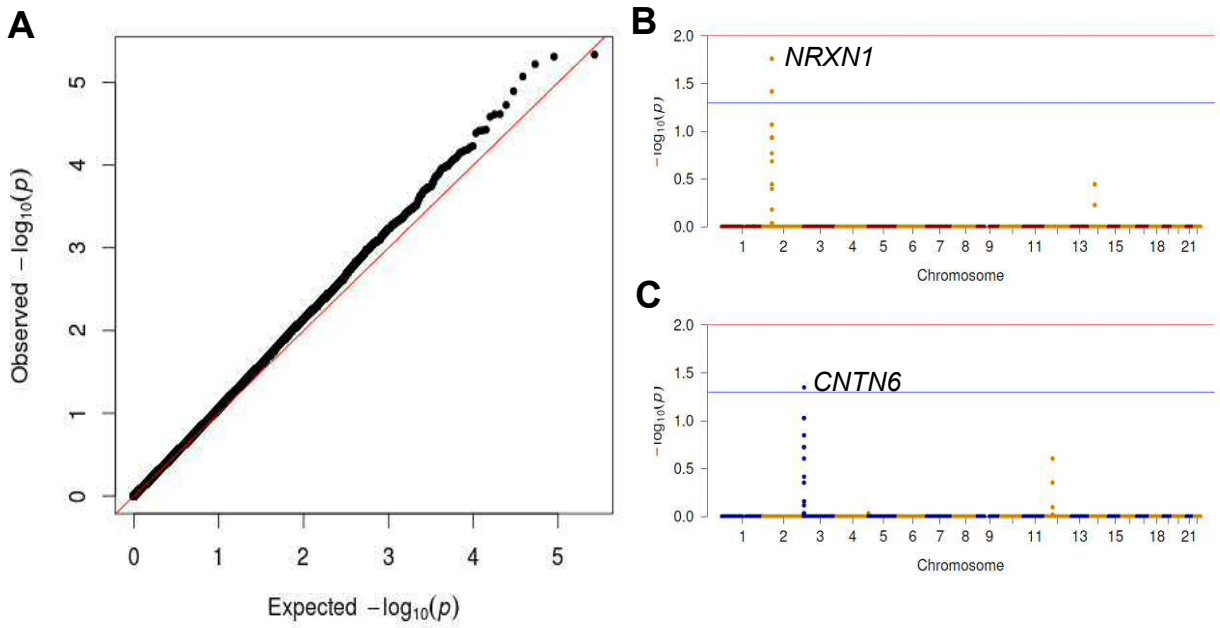


Figure S7. Examination of genome-wide TS case-control CNV analysis for population-specific effects (related to Figure 4). To verify the robustness of our results to population stratification, we pair-matched each case subject with exactly one control such that the global difference between all pairs is minimized using SpectralGEM. (A) The SNP-based λ_{gc} of the resultant dataset (1996 cases and 1996 controls) was an acceptable 1.082. Manhattan plots of segmental association results demonstrate that (B) deletions in *NRXN1* and (C) duplications in *CNTN6* are significant with an $\alpha < 0.05$ (blue line). Deletions and duplications were analyzed separately. The -log10 (p-value) displayed is empirically corrected for FWER (family-wise error rate) genome-wide using the max(T) method with 1,000,000 permutations.

Supplemental Tables

Table S1. Sample genotyping and QC summary (related to Figure 1). (A) Summary of included studies and genotyping information. Sample phenotypes, genotyping platform, and genotyping center for different datasets collected for this study are shown, separated by study. (B) Summary of quality control procedures by study. The number of samples remaining within each batch after each successive quality control step (see STAR Methods) is shown. Study abbreviations: Cardiff Controls (CC), Consortium for Neuropsychiatric Phenomics (CNP), Genomic Psychiatry Cohort (GPC), Wellcome Trust Case-Control Consortium (WTCCC2) and TS cases and controls collected for this study (GGRI1-3).

See attached tscnv_table-s1.docx

Table S2. Gaussian Mixture Model (GMM) clustered genotype calls at common Hapmap 3 CNVs (related to Figure 1, Figure S2, and Table S3). For sensitivity analysis, all 6,427 samples used in this study were genotyped across 11 common Hapmap3 CNVs using a locus-specific GMM-based clustering method (see Supplementary Methods). The numbers for CNV_ID: Hapmap3 accession number. CLUSTER_ID: Arbitrary identifier assigned by the clustering algorithm. CLUSTER_LRR: The mean value of all median-summarized intensity values for all samples assigned to the cluster. CLUSTER_COPY: Copy-number state inferred by examination of raw LRR-intensity values for samples within the cluster. Call frequencies (FREQ) for 4,093 controls (CTRL) and 2,434 TS cases (CASE) reflect the proportion of GMM-based genotype calls with >0.95 posterior probability of cluster assignment. There was no significant difference in CNV genotype frequency between phenotypic groups at any of the 21 non-reference genotype calls across all 11 loci (Fisher's exact test).

See attached tscnv_table-s2.docx

Table S3. Sensitivity analysis of consensus Hidden Markov Modeling (HMM) segmentation calls (related to Figure 1, Figure S2, and Table S2). (A) Sensitivity by locus. The sensitivity of HMM calling was defined as the number of concordant HMM calls divided by the total number of non-reference genotypes called in the same individual by Gaussian Mixture Model (GMM) clustering. Non-reference GMM calls were collapsed into calls of the same class (CNV_TYPE, DEL or DUP). (B) Overall sensitivity across all loci. P-values for both individual and aggregated locus-specific locus-based tests were calculated using Fisher's exact test. (C) Group-wise comparison of sensitivity between cases and controls based on the average sensitivity calculated within each individual. No significant difference in the average individual sensitivity was observed between phenotypic groups whether considering deletions, duplications, or both in concert (Welch's *t*-test).

See attached tscnv_table-s3.docx

Table S4. Rare CNV calls from 2,434 TS cases and 4,093 controls (related to Figures 1-4). This table contains a list of all rare CNVs detected in this study (STAR Methods). ID, sample ID; SEX, sample sex, Phenotype and comorbid disorders, TS, Tourette syndrome; OCD, obsessive-compulsive disorder; ADHD, attention deficit disorder; CHR, chromosome; START, CNV start (hg19 coordinates); STOP, CNV end; SIZE, CNV size in bp; TYPE, CNV copy-number; SITES, number of SNP markers spanning CNV; FREQUENCY, CNV frequency; RSID_START, first SNP within the CNV; RSID_STOP, last SNP within the CNV; GENE_COUNT, number of genes spanned by CNV; GENE_SYMBOLS, genes spanned by CNV. GENE_COUNT and GENE_SYMBOLS are determined by overlap with exons of protein-coding genes (STAR Methods).

See attached tscnv_table-s4.xlsx

Table S5. Clinical phenotypes of *NRXN1* and *CNTN6* CNV carriers. Clinical phenotypes for all CNV carriers of the two significant TS loci detected in this study: deletions at *NRXN1* and duplications at *CNTN6*. ID, Sample ID; Gene: Gene locus; Chr, chromosome; Start, CNV start (hg19 coordinates); End, CNV end; Type, CNV type; Length, CNV length (in kb); Variant Effect, CNV location within gene; Common comorbid disorders for TS, ADHD, attention deficit disorder; OCD, obsessive-compulsive disorder; Atypical, Case flagged for atypical presentation; Notes: Description of atypical phenotypic features. NA, No clinical information available.

See attached tscnv_table-s5.docx

Disclosures

P.S., M.G., H.S.S., R.A.K., Y.D., G.R., C.L.B., G.L., W.M.M., D.L.P, N.J.C., N.B.F., P.P., C.A.M. and J.M.S have received research funding from the Tourette Association of America (TAA). J.M.S., C.A.M. have received travel support from the TAA and serve on the TAA Scientific Advisory Board. P.S. received unrestricted Educational Grants in support of conferences he organized from Purdue and Shire, a CME speaker fee from Purdue University, industry sponsored clinical trial support from Otsuka and is a member of the Data Safety Monitoring Committee at Psyadon. C.L.B. has received funding for clinical trials from Psyadon Pharmaceuticals, Neurocrine Pharmaceuticals, Synchroneuron Pharmaceuticals, AUSPEX Pharmaceuticals, and TEVA Pharmaceuticals. She was a paid speaker for the TAA CDC program and a paid consultant for Bracket eCOA. I.A.M. has participated in research funded by the National Parkinson Foundation, TAA, Abbvie, Auspex, Biotie, Michael J. Fox Foundation, Neurocrine, Pfizer, and Teva, but has no owner interest in any pharmaceutical company. I.A.M. has been reimbursed for speaking for the National Parkinson Foundation and TAA. M.S.O. serves as a consultant for the National Parkinson Foundation, and has received research grants from NIH, NPF, the Michael J. Fox Foundation, the Parkinson Alliance, Smallwood Foundation, the Bachmann-Strauss Foundation, the TAA, and the UF Foundation. M.S.O's DBS research is supported by: R01 NR014852 and R01NS096008. He has previously received honoraria, but in the past >60 months has received no support from industry. M.S.O. has received royalties for publications with Demos, Manson, Amazon, Smashwords, Books4Patients, and Cambridge, is an associate editor for New England Journal of Medicine Journal Watch Neurology, has participated in CME and educational activities on movement disorders in the last 36 months sponsored by PeerView, Prime, QuantiaMD, WebMD, Medicus, MedNet, Henry Stewart, and by Vanderbilt University. The institution and not M.S.O. receives grants from Medtronic, Abbvie, Allergan, and ANS/St. Jude, and the PI has no financial interest in these grants. D.W.W. has

received royalties from Guilford Press and Oxford University Press, speaking honoraria from the TAA and serves on the TAA Medical Advisory Board. All other authors have no competing financial interests to declare.

None of the funding agencies for this project (NINDS, NIMH, the Tourette Association of America) had any influence or played any role in a) the design or conduct of the study; b) management, analysis, or interpretation of the data; or c) preparation, review, or approval of the manuscript.

Tourette Syndrome Association International Consortium for Genetics (TSAICG)

Cathy L. Barr[1], James R. Batterson[2], Cheston Berlin[3], Ruth D. Bruun[4], Cathy L. Budman[5], Danielle C. Cath[6], Sylvain Chouinard[7], Giovanni Coppola[8], Nancy J. Cox[9], Sabrina Darrow[10], Lea K. Davis[9], Yves Dion[11], Nelson B. Freimer[8], Marco A. Grados[12], Matthew E. Hirschtritt[10], Alden Y. Huang[8], Cornelia Illmann[13], Robert A. King[14], Roger Kurlan[15], James F. Leckman[14], Gholson J. Lyon[16], Irene A. Malaty[17], Carol A. Mathews[18], William M. MaMahon[19], Benjamin M. Neale[13], Michael S. Okun[17], Lisa Osiecki[13], David L. Pauls[13], Danielle Posthuma[20], Vasily Ramensky[8], Mary M. Robertson[21], Guy A. Rouleau[22], Paul Sandor[23], Jeremiah M. Scharf[13], Harvey S. Singer[12], Jan Smit[24], Jae-Hoon Sul[8], Dongmei Yu[13]

[1]Krembil Research Institute, University Health Network, Toronto, Ontario, Canada

[2]Children's Mercy Hospital, Kansas City, KS, USA

[3]Penn State University College of Medicine, Hershey, PA, USA

[4]Department of Psychiatry, North Shore-Long Island Jewish Medical Center, Manhasset, NY, USA

[5]Department of Psychiatry, North Shore University Hospital, Northwell Health System, Manhasset, NY, USA

[6]Department of Psychiatry, University Medical Center Groningen & Drenthe Mental Health Center, Netherlands

[7]Montreal Neurological Institute and University of Montreal, Montreal, Quebec, Canada

[8]Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

[9]Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

[10]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA

[11]Department of Psychiatry, University of Montreal, Montreal, Quebec, Canada

[12]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine
Baltimore, MD, USA

[13]Cornelia Illmann, Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[14]Yale Child Study Center, Yale University School of Medicine, New Haven, CT, USA

[15]Roger Kurlan, The Center for Neurological and Neurodevelopmental Health, Voorhees, NJ, USA

[16]Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

[17]Department of Neurology and Center for Movement Disorders and Neurorestoration, University of Florida, Gainesville, FL, USA

[18]Department of Psychiatry, and University of Florida Genetics Institute, University of Florida, Gainesville, FL, USA

[19]Department of Psychiatry, University of Utah, Salt Lake City, UT, USA

[20]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, VU University of Amsterdam, Amsterdam, Netherlands

[21]Division of Psychiatry, University College London, London, England

[22]Department of Neurology and Neurosurgery, Montreal Neurological Institute and McGill University
Montreal, Quebec, Canada

[23]Department of Psychiatry, University of Toronto and University Health Network, Youthdale Treatment Centers, Toronto, Ontario, Canada

[24]Free University of Amsterdam, University of Utrecht, Utrecht, Netherlands

The Gilles de la Tourette Syndrome Genome-wide Association Study Replication Initiative

Harald Aschauer[1], Csaba Barta[2], Danielle C. Cath[3,4], Christel Depienne[5,6], Andreas Hartmann[6], Johannes Hebebrand[7], Anastasios Konstantinidis[1], Kirsten R. Muller-Vahl[8], Peter Nagy[9], Markus M. Nöthen[10,11], Peristera Paschou[12,13], Renata Rizzo[14], Guy Rouleau[15], Monika Schlögelhofer[1], Mara Stamenkovic[1], Manfred Stuhrmann[16],Fotis Tsetsos[12,13], Zsanett Tarnok[9], Tomasz Wolanczyk[17], Yulia Worbe[6]

[1]Medical University Vienna, Department of Psychiatry and Psychotherapy, Vienna, Austria

[2]Institute of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary

[3]Department of Psychiatry, University Medical Center Groningen & Drenthe Mental Health Center, Groningen, the Netherlands

[4]Department of Clinical Psychology, Utrecht University, Utrecht, the Netherlands

[5]Département de Médicine Translationnelle et Neurogénétique, IGBMC, CNRS UMR 7104/INSERM U964/Université de Strasbourg,Illkirch, France

[6]Brain and Spine Institute, UPMC/INSERM UMR_S1127, Paris, France

[7]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

[8]Clinic of Psychiatry, Social Psychiatry and Psychotherapy, Hannover Medical School, Hannover, Germany

[9]Vadaskert Child and Adolescent Psychiatric Hospital, Budapest, Hungary

[10]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany

[11]Institute of Human Genetics, University of Bonn, Bonn, Germany

[12]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA

[13]Department of Molecular Biology and Genetics, Democritus University of Thrace, Greece

[14]Dipartimento di Medicina Clinica e Sperimentale, Università di Catania, Catania, Italy

[15]Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Canada

[16]Institute of Human Genetics, Hannover Medical School, Hannover, Germany

[17]Department of Child Psychiatry, Medical University of Warsaw, Warsaw, Poland

Table S4

Click here to access/download

Supplemental Movies & Spreadsheets

tscnv_table-s4.xlsx

STAR METHODS

Sample Ascertainment

Tourette Syndrome (TS) cases were ascertained primarily from TS specialty clinics through sites distributed throughout North America, Europe and Israel as part of an ongoing collaborative effort by the Tourette Syndrome Association International Consortium for Genetics (TSAICG) as described in detail elsewhere (Scharf et al., 2013). Subjects were assessed for a lifetime diagnosis of TS, Obsessive Compulsive Disorder (OCD) and Attention-Deficit/Hyperactivity Disorder (ADHD) using a standardized and validated semi-structured direct interview (TICS Inventory) (Darrow et al., 2015; Tourette Syndrome Association International Consortium for Genetics, 2007). An additional 628 cases were collected at 9 TS specialty clinics in Austria, Canada, France, Germany, Greece, Hungary, Italy and the Netherlands by expert clinicians using Tourette Syndrome Study Group criteria for Definite TS (DSM-IV TS diagnosis plus tics observed by a trained clinician) as well as DSM-IV diagnostic criteria for OCD and ADHD, along with 610 ancestry-matched controls as described previously (Paschou et al., 2014). We did not conduct formal standardized assessments for other neurodevelopmental disorders (NDDs) such as Intellectual Disability/Developmental Delay (ID/DD), Autism Spectrum Disorder (ASD), or schizophrenia/childhood psychosis; however, participants and their parents were asked about the presence of established or suspected diagnoses.

Additional TS case samples were obtained through web-based recruitment of individuals with a prior clinical diagnosis of TS who subsequently completed an online questionnaire that we have validated against the gold-standard TS structured diagnostic interview with nearly 100% concordance for all inclusion/exclusion criteria as well as high correct classification rates

for DSM-IV diagnoses of OCD and ADHD (Darrow et al., 2015; Egan et al., 2012). Individuals for web-based screening were solicited through the Tourette Association of America mailing list as well as from 4 TS specialty clinics in the United States. Individuals with a history of intellectual disability, seizure disorder, or a known tic disorder unrelated to TS were excluded.

External Control Sets

Additional control subjects were taken from four external large-scale genetic studies consisting of healthy individuals sampled from similar geographic locations, specifically selected because intensity data was available and generated on the same Illumina OmniExpress platform as the TS cases and controls collected as part of this study:

1. Cardiff Controls (CC): UK Blood donors were recruited in Cardiff at the time of blood donation at centres in Wales and England. Although not explicitly screened for psychiatric disorders, these controls are likely to have low rates of severe neuropsychiatric illness, as blood donors in the UK are only eligible to donate if they are not taking any medications. 57% of these controls were male. Genotyping on the Illumina OmniExpress was performed at
2. Consortium for Neuropsychiatric Phenomics (CNP): A collection of neuropsychiatric samples composed of patients with attention deficit hyperactivity disorder (ADHD), bipolar disorder (BD), schizophrenia (SCZ), and psychologically normal controls, collected throughout North America as part of a large NIH Roadmap interdisciplinary research consortia centered at the University of California, Los Angeles. Only the control samples were used in this study.

3. Genomic Psychiatry Cohort (GPC) (Pato et al., 2013): A large, longitudinal, population resource composed of clinically ascertained patients affected with BD, SCZ, their unaffected family members, and a large set of control samples with no family history of either disorder. Samples were collected at various sites throughout North America in a National Institute of Mental Health-sponsored study lead by the University of Southern California.
4. Wellcome Trust Case Control Consortium 2 (WTCCC2) (Power and Elliott, 2006): A subset of control samples from the National Blood Donors Cohort.

Genotyping and data preprocessing

Cases and controls collected specifically for this study (GGRI) were all genotyped on the Illumina OmniExpress Exome v1.1, while the remaining control samples from the CC, CNP, GPC, and WTCCC2 cohorts were genotyped on the Illumina OmniExpress 12v1.0. The content of these two arrays is identical except for 1) exome-focused content on the former and 2) additional intensity-only markers on the latter. We have observed that exome-specific assays in general exhibit a much higher variance overall in their derived log-R ratio (LRR). Therefore, in order to avoid detection biases due to this differential variance as well as due to unequal probe coverage, only the SNP assays with common identifiers across all versions of the OmniExpress arrays used in all datasets were used for quality control (QC) and CNV detection, a total of 689,077 markers.

To ensure the generation of the most reliable SNP calls, intensity measures, and B-allele frequencies (BAF) and to reduce the effect of differential processing, a custom cluster file was generated for each dataset separately, and each genotype batch when such information was

available. Since the performance of Illumina's proprietary normalization and cluster generation process improves with the number of samples, we processed all of the raw intensity data available, regardless of clinical phenotype phenotype. An initial round of QC was carried out using Illumina Beeline to determine baseline calling rates for each sample using the canonical cluster file (*.egt) provided by the manufacturer for each array version. Any sample with a call rate < 0.98 or a log-R ratio (LRR) standard deviation > 0.30 was deemed a failed assay and removed (Pre-cluster QC, Supplementary Table 1B). SNP clustering and genotype calling was then performed in GenomeStudio with only passing samples within each dataset.

Genotype sample QC

We performed an initial round of QC based on SNP genotype data. All samples at this stage had a minimal call rate >0.98. For duplicate samples, we retained samples with the higher call rate. Additionally, we removed samples with with discordant sex status and excess autosomal heterozygosity rate (total sample mean +/- 1.5SD). We filtered autosomal SNPs for missingness > 0.01; minor-allele frequency > 0.01; deviation from Hardy-Weinberg equilibrium (P < 10^-6); and pruned SNPs for LD (r^2 < 0.05). We used a total of 62,104 SNPs for relatedness testing using PLINK. For all pairs of subjects with PI_HAT > 0.185, we removed one subject at random, with the exception that if a control individual was related to a subject with a neuropsychiatric phenotype, it was explicitly marked for removal.

Genome-wide detection of CNV loci

We employed two widely-used HMM-based CNV calling algorithms, PennCNV (Wang et al., 2007) (version 2011-05-03, http://penncnv.openbioinformatics.org/en/latest/), and QuantiSNP (Colella et al., 2007) (version 2.0, https://sites.google.com/site/quantisnp/downloads), to initially detect structural variants in our dataset. We created GC-adjusted LRR intensity files for all samples using the GC-waviness correction method (Diskin et al., 2008). For PennCNV, a custom population B-allele frequency file was created for each dataset separately and CNV calls were generated using the standard protocol. QuantiSNP calls were similarly generated on the GC-adjusted intensity files. Using custom PERL scripts, a concordant callset between both CNV callers was then generated by taking the intersecting boundaries of overlapping calls of the same CNV type (deletion or duplication). Additionally, adjacent CNV calls were merged if they were spanned by a CNV called by the other HMM algorithm. As HMMs have been shown to artificially break up large CNVs, we also merged CNV segments in the final concordant callset if they were of the same copy number and the number of intervening markers between them was less than 20% of the total of both segments combined using the PennCNV's "clean_cnv.pl" script. We repeated this joining process iteratively until no more merging of segments occurred.

Array intensity sample QC

The PennCNV calling algorithm generates a number of array intensity-based metrics with regard to CNV assay quality. Intensity-based QC was conducted based on the distribution of all available assays and subsequently combined with the results from the SNP-based QC. To

remove samples with data unsuited for CNV detection, we used empirically defined thresholds across several different metrics:

1. Waviness factor (WF) - measures the waviness in intensity values, a known artifact caused by improper DNA concentration that can lead to spurious calls.
2. Log-R ratio standard deviation (LRR-SD) - a measure of the overall variance in intensity values.
3. B-allele frequency drift standard deviation (BAF-SD) - the standard deviation of autosomal BAF values between 0.25 and 0.75.

Thresholds for WF and LRR-SD were determined separately for each dataset by both manual examination of QC metrics and taking the mean + 3 x SD to determine outlying samples. As the LRR is essentially composed of the sum of allelic intensities, it is highly correlated with BAF-SD. Therefore, instead of a hard cutoff, we opted to remove outliers in BAF-SD while considering LRRSD. We plotted sample BAF-SD values ranked by LRR-SD and fit a loess curve using the R function *loess.smooth*, adjusting the span parameter until the distribution was adequately fit. We then shifted the fitted curve by an appropriate factor (typically about 1.2) to establish a cutoff for BAF-SD. This was done for each dataset separately. Following intensity-based QC, all samples had an LRR-SD of <0.24, absolute value of WF < 0.04, and an BAF-SD < 0.05, well within established limits required for reliable CNV detection.

EU-ancestry estimation

Following sample removal based on SNP and intensity-based QC, we removed all clinical non-TS samples from external studies, and genotype data for all remaining samples was

subsequently combined with data from publicly available HapMap samples of European, African, and Asian continental ancestry (Illumina). All available European (EU) population samples from the 1000 Genomes Project were also included to establish an appropriate calibration threshold for EU ancestry designation. A randomly-thinned dataset comprised of 19,024 LD-independent markers were used for ancestry inference using fastStructure with k=3, and samples were excluded if they contained > 0.0985 non-EU ancestry (Supplementary Figure 1). A final round of ancestry exclusion was performed by removing all samples outside of the median +/- 3SD of the first three ancestry PCs.

CNV load sample QC

Although SNP and intensity-based QC removed most failed assays, we performed a final round of sample QC, removing eight additional samples with excessively high CNV load based on the total number of CNV calls (>45) or total CNV length (>10Mb). These thresholds were determined empirically by visual inspection of distributions across all datasets combined. Our final dataset after QC consisted 6,527 samples: 2,434 TS cases and 4,093 controls.

Data handling and CNV visualization

To facilitate further data processing and visualization of CNV events, we generated an HDF5 database consisting of sample metadata, CNV calls, probe information, LRR intensity and BAF values for all assays. Normalized intensity values were also generated by converting the GC-corrected, median-centered LRR measures into Z-scores within each sample and inserted into the database. Calculations for the sensitivity assessment (following section) were

performed in Python. Visualization of GMM cluster plots (Figures S2A, S2B, S3C); median Z-score outlier detection (MeZOD) CNV calls (Kirov et al., 2012) (Figure 3B); and probe-level CNV plots (Figures S5 and S6) were generated from Z-scores of intensity data using in Matplotlib.

Sensitivity assessment

We augmented a previously described method (Vacic et al., 2011) to investigate whether any difference in sensitivity to detect CNVs existed between cases and controls within the context of our study. Both HMM-based CNV callers we employed for genome-wide detection are univariate methods completely agnostic of intensity information across multiple samples and do not use known population frequency prior probabilities in their calling algorithms. Therefore, common CNVs act as an ideal proxy to evaluate the effectiveness to detect rare events accurately; they are detected in the same manner but are present at much higher frequencies, enabling an accurate estimation of the overall sensitivity of detection for rare events genome-wide.

We used the UCSC Genome Browser (https://genome.ucsc.edu) liftOver tool to translate a list of common HapMap3 CNVs to the hg19 reference. To match the thresholds used for our association tests in this study, we filtered the list of common CNVs to those that were >30 kbp in length. We reduced the number of markers required slightly to a minimum of 9 to ensure that an adequate number of events could be assessed. For each common CNV meeting these criteria, we examined the distribution of median-summarized normalized intensity measures within the CNV region across all study samples and retained only those loci that displayed no evidence of

clustering intro different copy-number states. A total of 11 common CNV loci were retained for sensitivity analysis (Figure S2 and Table S2).

We generated locus-specific genotyping calls in the following manner. First, we extracted the LRR intensity Z-scores for all probes in the region across all samples. The Z-scores for all probes spanning the CNV locus were then subjected to second round of normalization across all samples. A Gaussian mixture model (GMM) was fit to this distribution of Z-scores using the SciKit-learn Python package. The optimum number of clusters was automatically determined by minimization of the Bayesian Information Criterion (BIC) and corrected, when necessary, by manual adjustment. Individuals were assigned to a cluster only if the posterior probability of assignment exceeded 0.95.

Copy number state was inferred by examining the original LRR intensity values for samples within each cluster. We inspected for allele frequency differences between controls and cases for all clusters and found no significant difference (Fisher Exact Test, Table S2). We collapsed the clusters at each locus into CNVs of the same type (deletion or duplication), and, as this locus-specific genotyping method is more sensitive than HMM segmentation methods, we used the proportion of concordant of HMM-based calls as a measure of segmentation detection sensitivity. We found no significant difference in sensitivity to detect common CNVs between phenotypic groups at any of the 11 loci tested, either independently, or in concert (Fisher Exact Test, Supplementary Table S3A and S3B). Furthermore, the mean sensitivity for each sample was calculated and collectively assessed for any systematic difference between phenotypic groups. Considering duplications, deletions, or both in concert, we observed no significant difference in the sensitivity of segmentation calls between case and control groups (Wald T-test, Supplementary Table S3C).

Determination of rare CNVs

Calls were removed from the dataset if they spanned less than 10 markers, were less than 30kb in length, or overlapped by more than 0.5 of their total length with regions known to generate artifacts in SNP-based detection of CNVs. This included immunoglobulin domain regions, segmental duplications, and regions that have previously demonstrated associations specific to Epstein-Barr virus immortalized cell lines (Shirley et al., 2012). In addition, we removed calls that spanned telomeric (defined as 100kb from the chromosome ends), centromeric regions, and gaps in the reference genome. We assigned all CNV calls a specific frequency count using the PLINK command `--cnv-freq-method2 0.5`. Here, the frequency count of an individual CNV is determined as 1 + the total number of CNVs overlap by at least 50% of its total length (in bp), irrespective of CNV type. We then filtered our callset for CNVs with MAF < 1% (a frequency of 65 or lower across 6,527 samples). Furthermore, we removed calls if they overlapped by more than 50% of their length to common CNVs regions determined by several high-quality, publically available SNP-array datasets:

1. 845 population samples from the Deciphering Developmental Disorders study (compiled by DECIPHER[1])
2. 450 population samples from the 42M genotyped study (compiled by DECIPHER)
3. 5919 population samples from the Affy6 study (compiled by DECIPHER)

[1] https://decipher.sanger.ac.uk/index

Conservatively, CNVs regions were considered common if present at a frequency of >10% within any individual dataset above.

In-silico validation

For each putative rare CNV, we generated two different metrics based on intensity and B-allele frequency (BAF) banding. To qualifying CNVs based on intensity, we adopted a methodology similar to the MeZOD method described elsewhere (Kirov et al., 2012), with modification. We observed that standardized intensity measures typically range from < -20 for homozygous deletions, [-6,-2.3] for heterozygous deletions, and > 1.3 for duplications. Because of the disproportionately large effect on intensity measures caused deletions events, performing a second round of normalization across all samples within each putative CNV will skew the overall distribution when homozygous deletion events are present. Therefore, we only performed a single round of normalization of LRR intensity measures, within each sample. Each CNV was scored by calculating the median of LRR intensity Z-scores (LRR-Z) for all probes within the region. To qualify CNVs based on BAF banding, we calculated the proportion of probes within the CNV region that showed evidence of a duplication event (BAF of [0.25-04] or [0.6-0.8]), and denoted this measure “BAF-D.”

Based on examination of our own data, as prior observations (Sanders et al., 2015) we flagged deletions that had a LRR-Z > -2 or BAF-D > 0.02, and duplications with a LRR-Z < 1 or BAF-D < 0.1. To avoid differential missingness caused by subtle differences between arrays, we did not impose any hard cutoff for CNV exclusion based on these metrics. These thresholds were applied to flag CNV calls with marginal scores for manual inspection and the removal

obviously misclassified events. Through this in-silico validation process we discovered multiple instances of CNV calls likely due to individual mosaicism (Figure S3A and S3B), and removed these events from subsequent analysis.

Furthermore, for each rare CNV call, we used distribution of summarized intensity information across all individuals For each rare CNV region, we quantified proportion of samples whose LRR-Z metric fell outside of [-2.3, 1.3] and further inspected these regions manually. Putative rare CNV loci that showed substantial evidence for extensive polymorphism were subsequently scored for frequency using the GMM genotyping method described above (Figure S3C) and conservatively, only removed if shown to be variant in more than 10% of the samples across the entire dataset.

Annotation of rare cnvs

Rare CNVs were annotated for gene content according to RefSeq provided by the hg19 assembly of the UCSC Genome Browser. We only considered a CNV as “genic” if it overlapped any exon of a known protein-coding transcript (as designated by the RefSeq transcript accession prefix “NM”). The “gene count” of a CNV represents the total number of non-redundant genes whose respective transcripts it overlaps. “Non-genic CNVs” represent all variants that are not genic according to the definition above. In addition, all rare CNVs were assessed for clinical relevance in accordance with guidelines set forth by the American College of Medical Genetics (ACMG). This was accomplished through the use of the Scripps Genome Advisor (SGA) (https://genomics.scripps.edu/ADVISER/); inspection of curated resources including ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), DECIPHER (https://decipher.sanger.ac.uk/index), and Online Mendelian Inheritance in Man

(https://www.omim.org/); and followed by confirmation through review of primary literature. Conservatively, we only considered a CNV as "pathogenic" if directly supported by more than one primary publication. To screen for additional relevant pathogenic variants, we screened for variants that overlapped compiled lists defined in the literature (Malhotra and Sebat, 2012; McGrath et al., 2014). All non-pathogenic variants were automatically annotated by SGA: those classed as Category 2-4 were assigned as a variant of "unknown clinical significance", and those assigned to Category 5 were considered "benign". Note that, as only rare CNVs were considered, the number of "benign" variants is small.

Global CNV burden analysis

We measured global CNV burden using three separate metrics: the number of rare CNVs (CNV count), the total length of all CNVs (CNV length) and the total number of genes intersected by CNVs (CNV gene count). To examine the effect of different covariates on our different metrics of global CNV burden, we first fit a linear regression model for each burden metric, using the *glm* function in R:

Burden_metric ~ TS_status + subject_sex + LRR_SD + ancestry_PCs

In the above model, as assay intensity quality metrics are highly correlated, we utilized LRR_SD as a single measure of assay quality, and ancestry_PCs include the first ten principal components (PCs) of ancestry derived from SNP data. Of the above included covariates (excluding TS_status), under any metric considered, only LRR_SD, PC1, and PC2 were significant predictors of global CNV burden ($P<0.05$). To assess for a global burden difference between TS cases and controls, we fit a logistic regression model in R:

TS_status ~ Burden_metric + subject_sex + LRR_SD + PC1 + PC2 + PC3 + PC4

Independent variables included burden metric, subject sex, LRR_SD, and, to correct for potential population stratification associated with the case-control status, the first four ancestry PCs. We note that our logistic model includes all covariates that were found to be significantly associated with global CNV burden as determined by linear regression described above. ORs above 1 indicate an increased risk for TS per unit of CNV burden and P-values were calculated using the likelihood ratio test. In Figure 2, global burden was analyzed separately for genic CNVs and non-genic CNVs. Since we were interested in comparing the relative contribution to TS risk by different measures of burden and across various categories, sizes, and frequency classes, the ORs presented in Figure 2 and Figure S4 are calculated from standardized CNV burden metrics. In Figure 3, we calculated OR using the unstandardized value of actual CNV counts, as this is directly interpretable.

CNV burden of evolutionarily conserved genes

Rare genic CNVs were annotated using pLI (probability of LoF Intolerance) scores (http://exac.broadinstitute.org/). A CNV was marked as “constrained” if it overlapped any exon of a gene with a pLI score of > 0.9, as per the approach by Ruderfer and colleagues (Ruderfer et al., 2016). To examine whether such constrained CNVs are enriched in TS patients, we assumed a Poisson distribution of such rare events, and conducted a Poisson regression with adjustment of subject sex, LRR_SD, the first four ancestry PCs:

Constrained_CNVs ~ TS_status + subject_sex + LRR_SD + PC1 + PC2 + PC3 + PC4

We tested the large (>1Mb), singleton CNVs that affect conserved genes (pLI>0.9) stratified by CNV type, and compared the relative risk of such events to all large, singleton genic CNVs.

Locus-specific tests of association

The segmental test of association was performed by quantifying the frequencies of case and control CNV carriers at all unique CNV breakpoint locations; the unique set of CNV breakpoints defines all locations genome-wide where the frequency of CNVs can differ between cases and controls. For gene-based association tests, we restricted our analysis to genic CNVs (CNVs that intersect an exon of any protein-coding transcript, as defined above) and quantified the frequencies of cases and control CNVs across each gene. Locus-specific p-values for both tests of association were determined by 1,000,000 permutations of phenotype labels, and genome-wide corrected p-values were obtained using the max(T) permutation method as implemented in PLINK v1.07, which controls for family-wise error rate by comparing the locus-specific test statistic to all test statistics genome-wide within each permutation. Association tests were conducted separately for deletions and duplications.

Sensitivity analysis of association results

The segmental association test after carefully pair-matching each case with a control such that the global difference between each pair was minimized using SpectralGEM (Lee et al., 2010) (Figure S6). For the matched segmental association analysis, because of the drastic reduction in sample size, a genome-wide corrected alpha < 0.05 was used as a cutoff to indicate significance.

WORKS CITED

Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C., and Ragoussis, J. (2007). QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. *35*, 2013–2025.

Darrow, S.M., Illmann, C., Gauvin, C., Osiecki, L., Egan, C.A., Greenberg, E., Eckfield, M., Hirschtritt, M.E., Pauls, D.L., Batterson, J.R., et al. (2015). Web-based phenotyping for Tourette Syndrome: Reliability of common co-morbid diagnoses. Psychiatry Res. *228*, 816–825.

Diskin, S.J., Li, M., Hou, C., Yang, S., Glessner, J., Hakonarson, H., Bucan, M., Maris, J.M., and Wang, K. (2008). Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. *36*, e126.

Egan, C.A., Marakovitz, S.E., O'Rourke, J.A., Osiecki, L., Illmann, C., Barton, L., McLaughlin, E., Proujansky, R., Royal, J., Cowley, H., et al. (2012). Effectiveness of a web-based protocol for the screening and phenotyping of individuals with Tourette syndrome for genetic studies. Am. J. Med. Genet. B Neuropsychiatr. Genet. *159B*, 987–996.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry *17*, 142–153.

Lee, A.B., Luca, D., Klei, L., Devlin, B., and Roeder, K. (2010). Discovering genetic ancestry using spectral graph theory. Genet. Epidemiol. *34*, 51–59.

Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell *148*, 1223–1241.

McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E., Kusenda, M., Krastoshevsky, O., et al. (2009). Microduplications of 16p11.2 are associated with schizophrenia. Nat. Genet. *41*, 1223–1227.

McGrath, L.M., Yu, D., Marshall, C., Davis, L.K., Thiruvahindrapuram, B., Li, B., Cappi, C., Gerber, G., Wolf, A., Schroeder, F.A., et al. (2014). Copy number variation in obsessive-compulsive disorder and tourette syndrome: a cross-disorder study. J. Am. Acad. Child Adolesc. Psychiatry *53*, 910–919.

Paschou, P., Yu, D., Gerber, G., Evans, P., Tsetsos, F., Davis, L.K., Karagiannidis, I., Chaponis, J., Gamazon, E., Mueller-Vahl, K., et al. (2014). Genetic association signal near NTN4 in Tourette syndrome. Ann. Neurol. *76*, 310–315.

Pato, M.T., Sobell, J.L., Medeiros, H., Abbott, C., Sklar, B.M., Buckley, P.F., Bromet, E.J., Escamilla, M.A., Fanous, A.H., Lehrer, D.S., et al. (2013). The genomic psychiatry cohort: partners in discovery. Am. J. Med. Genet. B Neuropsychiatr. Genet. *162B*, 306–312.

Power, C., and Elliott, J. (2006). Cohort profile: 1958 British birth cohort (National Child

Development Study). Int. J. Epidemiol. *35*, 34–41.

Ruderfer, D.M., Hamamsy, T., Lek, M., Karczewski, K.J., Kavanagh, D., Samocha, K.E., Exome Aggregation Consortium, Daly, M.J., MacArthur, D.G., Fromer, M., et al. (2016). Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. Nat. Genet.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron *87*, 1215–1233.

Scharf, J.M., Yu, D., Mathews, C.A., Neale, B.M., Stewart, S.E., Fagerness, J.A., Evans, P., Gamazon, E., Edlund, C.K., Service, S.K., et al. (2013). Genome-wide association study of Tourette's syndrome. Mol. Psychiatry *18*, 721–728.

Shirley, M.D., Baugher, J.D., Stevens, E.L., Tang, Z., Gerry, N., Beiswanger, C.M., Berlin, D.S., and Pevsner, J. (2012). Chromosomal variation in lymphoblastoid cell lines. Hum. Mutat. *33*, 1075–1086.

Tourette Syndrome Association International Consortium for Genetics (2007). Genome scan for Tourette disorder in affected-sibling-pair and multigenerational families. Am. J. Hum. Genet. *80*, 265–272.

Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H.-H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R., et al. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature *471*, 499–503.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. *17*, 1665–1674.